

Learning by Approximation

J. Kostlivá, Z. Straka, P. Švarný

Today two examples:

1. Approximation in least square sense
2. Approximative Q-learning

Least square approximation

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How?:

- A: minimize difference in coordinates
- B: maximize error
- C: minimize sum of squared errors
- D: maximize difference in coordinates

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How?:

A: minimize difference in coordinates

B: maximize error

C: minimize sum of squared errors

D: maximize difference in coordinates

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How?:

- A: minimize difference in coordinates
- B: maximize error
- C: minimize sum of squared errors
- D: maximize difference in coordinates

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How?:

- A: minimize difference in coordinates
- B: maximize error
- C: minimize sum of squared errors
- D: maximize difference in coordinates

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How? - minimize sum of squared errors.

Define:

$$A: \sum_i (f(x_i) - x_i)^2$$

$$B: \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$$

$$C: \sum_i (x_i - f(x_i))^2$$

$$D: \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How? - minimize sum of squared errors.

Define:

A: $\sum_i (f(x_i) - x_i)^2$

B: $\sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

C: $\sum_i (x_i - f(x_i))^2$

D: $\sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

How? - minimize sum of squared errors.

Define:

A: $\sum_i (f(x_i) - x_i)^2$

B: $\sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

C: $\sum_i (x_i - f(x_i))^2$

D: $\sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

How?:

A: find solution of $E = 0$

B: find maximum of E

C: find minimum of E

D: find solution $E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

How?:

- A: find solution of $E = 0$
- B: find maximum of E
- C: find minimum of E
- D: find solution $E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

How?:

- A: find solution of $E = 0$
- B: find maximum of E
- C: find minimum of E
- D: find solution $E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E .

How? Solve:

A: $E = 0$

B: $\partial E = 0$

C: $E = -\infty$

D: $\partial E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E .

How? Solve:

A: $E = 0$

B: $\partial E = 0$

C: $E = -\infty$

D: $\partial E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E .

How? Solve:

A: $E = 0$

B: $\partial E = 0$

C: $E = -\infty$

D: $\partial E = -\infty$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\partial E = 0$

Derive by:

A: x

B: \mathbf{w}

C: w_1

D: $f(x_i)$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\partial E = 0$

Derive by:

- A: x
- B: \mathbf{w}
- C: w_1
- D: $f(x_i)$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\partial E = 0$

Derive by:

A: x

B: \mathbf{w}

C: w_1

D: $f(x_i)$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

Evaluate $\frac{\partial E}{\partial w_0}$:

A: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i - f(x_i))$

B: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + 1 - f(x_i))$

C: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_0} = 2 \sum_i (x_i - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

Evaluate $\frac{\partial E}{\partial w_0}$:

A: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i - f(x_i))$

B: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + 1 - f(x_i))$

C: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_0} = 2 \sum_i (x_i - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

Evaluate $\frac{\partial E}{\partial w_0}$:

A: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i - f(x_i))$

B: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + 1 - f(x_i))$

C: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_0} = 2 \sum_i (x_i - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

- ▶ $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

Evaluate $\frac{\partial E}{\partial w_1}$:

A: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i - f(x_i))x_i$

B: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i$

C: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1 + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_1} = 2 \sum_i (x_i + w_0 - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

- ▶ $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

Evaluate $\frac{\partial E}{\partial w_1}$:

A: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i - f(x_i))x_i$

B: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i$

C: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1 + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_1} = 2 \sum_i (x_i + w_0 - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

- ▶ $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i))$

Evaluate $\frac{\partial E}{\partial w_1}$:

A: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i - f(x_i))x_i$

B: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i$

C: $\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1 + w_0 - f(x_i))$

D: $\frac{\partial E}{\partial w_1} = 2 \sum_i (x_i + w_0 - f(x_i))$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Solve linear equation system.

Using given tuples (for simplicity let's use only first three tuples).

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Solve linear equation system.

Using given tuples (for simplicity let's use only first three tuples).

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Solve linear equation system.

Using given tuples (for simplicity let's use only first three tuples).

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = 2 \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Solve linear equation system.

Using given tuples (for simplicity let's use only first three tuples).

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

A: $\frac{\partial E}{\partial w_0} = w_1 - w_0 + 5$

B: $\frac{\partial E}{\partial w_0} = 2w_1 + w_0 - 4.2$

C: $\frac{\partial E}{\partial w_0} = 3w_1 + 3w_0 - 10.6$

D: $\frac{\partial E}{\partial w_0} = w_1 - 2w_0 - 3.1$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

A: $\frac{\partial E}{\partial w_0} = w_1 - w_0 + 5$

B: $\frac{\partial E}{\partial w_0} = 2w_1 + w_0 - 4.2$

C: $\frac{\partial E}{\partial w_0} = 3w_1 + 3w_0 - 10.6$

D: $\frac{\partial E}{\partial w_0} = w_1 - 2w_0 - 3.1$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

A: $\frac{\partial E}{\partial w_0} = w_1 - w_0 + 5$

B: $\frac{\partial E}{\partial w_0} = 2w_1 + w_0 - 4.2$

C: $\frac{\partial E}{\partial w_0} = (w_1 \cdot 0 + w_0 - 2.1) + (w_1 \cdot 1 + w_0 - 3.6) + (w_1 \cdot 2 + w_0 - 4.9) = 3w_1 + 3w_0 - 10.6$

D: $\frac{\partial E}{\partial w_0} = w_1 - 2w_0 - 3.1$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

A: $\frac{\partial E}{\partial w_1} = 5w_1 + 3w_0 - 13.4$

B: $\frac{\partial E}{\partial w_1} = 2w_1 + 6.2$

C: $\frac{\partial E}{\partial w_1} = w_1 + w_0 - 2.4$

D: $\frac{\partial E}{\partial w_1} = 2w_0 - 3.1$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

A: $\frac{\partial E}{\partial w_1} = 5w_1 + 3w_0 - 13.4$

B: $\frac{\partial E}{\partial w_1} = 2w_1 + 6.2$

C: $\frac{\partial E}{\partial w_1} = w_1 + w_0 - 2.4$

D: $\frac{\partial E}{\partial w_1} = 2w_0 - 3.1$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 0$$

Evaluate:

$$\text{A: } \frac{\partial E}{\partial w_1} = (w_1 \cdot 0 + w_0 - 2.1) \cdot 0 + (w_1 \cdot 1 + w_0 - 3.6) \cdot 1 + (w_1 \cdot 2 + w_0 - 4.9) \cdot 2 = 5w_1 + 3w_0 - 13.4$$

$$\text{B: } \frac{\partial E}{\partial w_1} = 2w_1 + 6.2$$

$$\text{C: } \frac{\partial E}{\partial w_1} = w_1 + w_0 - 2.4$$

$$\text{D: } \frac{\partial E}{\partial w_1} = 2w_0 - 3.1$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 5w_1 + 3w_0 - 13.4 = 0 \quad / \cdot (-1)$$

$$-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$$

$$w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$$

$$\Rightarrow \hat{f}(x, \mathbf{w}) = 1.4x + 2.133$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 5w_1 + 3w_0 - 13.4 = 0 \quad / \cdot (-1)$$

$$-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$$

$$w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$$

$$\Rightarrow \hat{f}(x, \mathbf{w}) = 1.4x + 2.133$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 5w_1 + 3w_0 - 13.4 = 0 \quad / \cdot (-1)$$

$$-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$$

$$w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$$

$$\Rightarrow \hat{f}(x, \mathbf{w}) = 1.4x + 2.133$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 5w_1 + 3w_0 - 13.4 = 0 \quad / \cdot (-1)$$

$$-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$$

$$w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$$

$$\Rightarrow \hat{f}(x, \mathbf{w}) = 1.4x + 2.133$$

Least square approximation

We have:

- ▶ given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

Task: determine/compute parameters w_0, w_1 with lowest error

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$

Find minimum of E by derivation $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1x_i + w_0 - f(x_i))^2 = 0$

$$\frac{\partial E}{\partial w_0} = \sum_i (w_1x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$$

$$\frac{\partial E}{\partial w_1} = \sum_i (w_1x_i + w_0 - f(x_i))x_i = 5w_1 + 3w_0 - 13.4 = 0 \quad / \cdot (-1)$$

$$-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$$

$$w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$$

$$\Rightarrow \hat{f}(x, \mathbf{w}) = 1.4x + 2.133$$

Approximative Q-learning

Approximative Q-learning

We have:

- ▶ an unknown grid world
- ▶ a few episodes the robot tried

Today:

- ▶ we approximate Q-function
- ▶ $\hat{q}(s, a, \mathbf{w}) = a w_1 + (1 - a) w_0$
- ▶ we will compute parameters w_0, w_1

Approximative Q-learning

We have:

- ▶ an unknown grid world
- ▶ a few episodes the robot tried

Today:

- ▶ we approximate Q-function
- ▶ $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$
- ▶ we will compute parameters w_0, w_1

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = a w_1 + (1 - a) w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = \alpha w_1 + (1 - \alpha) w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = \alpha w_1 + (1 - \alpha) w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

SGD briefly:

- ▶ Find \mathbf{w} that minimize $\sum_t (\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$
- ▶ How to do it online?
- ▶ In every timestep t , modify \mathbf{w} that value of $(\text{trial}_t - \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease.
- ▶ How?

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

How?:

A: $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\hat{q}(s_t, a_t, \mathbf{w}) + \alpha(\hat{q}(s_t, a_t, \mathbf{w}))$

B: $\hat{q}(s_t, a_t, \mathbf{w}) \leftarrow \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))$

C: $\hat{q}(s_t, a_t, \mathbf{w}) \leftarrow \hat{q}(s_t, a_t, \mathbf{w}) + \alpha(\text{trial})$

D: $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

How?:

A: $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\hat{q}(s_t, a_t, \mathbf{w}) + \alpha(\hat{q}(s_t, a_t, \mathbf{w}))$

B: $\hat{q}(s_t, a_t, \mathbf{w}) \leftarrow \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))$

C: $\hat{q}(s_t, a_t, \mathbf{w}) \leftarrow \hat{q}(s_t, a_t, \mathbf{w}) + \alpha(\text{trial})$

D: $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w})$$

Define:

$$\text{A: trial} = r_{t+1} + \gamma \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{B: trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{C: trial} = \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{D: trial} = r_{t+1} + \gamma \max_a \hat{q}(s_t, a, \mathbf{w})$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w})$$

Define:

$$\text{A: trial} = r_{t+1} + \gamma \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{B: trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{C: trial} = \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$\text{D: trial} = r_{t+1} + \gamma \max_a \hat{q}(s_t, a, \mathbf{w})$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$$

Define:

$$A: \text{trial} = r_{t+1} + \gamma\hat{q}(s_{t+1}, a, \mathbf{w})$$

$$B: \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$C: \text{trial} = \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$D: \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_t, a, \mathbf{w})$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Define w_1 update:

$$A: w_1^{t+1} = w_1^t + \alpha(\hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$B: w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$$

$$C: w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$D: w_1^{t+1} = w_1^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

Define w_1 update:

$$\text{A: } w_1^{t+1} = w_1^t + \alpha(\hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$\text{B: } w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$$

$$\text{C: } w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$\text{D: } w_1^{t+1} = w_1^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Define w_1 update:

A: $w_1^{t+1} = w_1^t + \alpha(\hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$

B: $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$

C: $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$

D: $w_1^{t+1} = w_1^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
 $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Define w_0 update:

$$\text{A: } w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$$

$$\text{B: } w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\text{C: } w_0^{t+1} = w_0^t + \alpha(\text{trial})(1 - a_t)$$

$$\text{D: } w_0^{t+1} = w_0^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
 $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Define w_0 update:

- A: $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$
- B: $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
- C: $w_0^{t+1} = w_0^t + \alpha(\text{trial})(1 - a_t)$
- D: $w_0^{t+1} = w_0^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
 $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Define w_0 update:

- A: $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$
- B: $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
- C: $w_0^{t+1} = w_0^t + \alpha(\text{trial})(1 - a_t)$
- D: $w_0^{t+1} = w_0^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

Let's compute $\mathbf{w} = (w_1, w_0)$

For simplicity: $\gamma = 1, \alpha = 1$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

Let's compute $\mathbf{w} = (w_1, w_0)$

For simplicity: $\gamma = 1, \alpha = 1$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

Let's compute $\mathbf{w} = (w_1, w_0)$

For simplicity: $\gamma = 1, \alpha = 1$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
 $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Initialize \mathbf{w} :

A: $\mathbf{w} = (w_1, w_0) = (1, 1)$

B: $\mathbf{w} = (w_1, w_0) = (0, 1)$

C: $\mathbf{w} = (w_1, w_0) = (0, 0)$

D: arbitrarily

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

- ▶ $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$
 $w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
- ▶ $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

Initialize \mathbf{w} :

A: $\mathbf{w} = (w_1, w_0) = (1, 1)$

B: $\mathbf{w} = (w_1, w_0) = (0, 1)$

C: $\mathbf{w} = (w_1, w_0) = (0, 0)$

D: arbitrarily (we choose $\mathbf{w} = (w_1, w_0) = (0, 0)$)

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1:$

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 1

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 1

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla\hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 1$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}))\nabla \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1$:

Compute:

$$\text{A: trial} = -2 + \max\{\hat{q}(s_{t+1} = 1, a = 0, \mathbf{w}^t), \hat{q}(s_{t+1} = 1, a = 1, \mathbf{w}^t)\} = -2 + \max\{0, 0\} = -2$$

$$\text{B: trial} = 0$$

$$\text{C: trial} = -1$$

$$\text{D: trial} = 1$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1$: trial = -2

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 1

C: diff = -1

D: diff = -2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2), t = 1$: trial = -2

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 1

C: diff = -1

D: diff = $-2 - 0 = -2$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2)$, $t = 1$: trial = -2, diff = -2

Compute :

$$A: w_1^{t+1} = 2$$

$$B: w_1^{t+1} = 0$$

$$C: w_1^{t+1} = 1$$

$$D: w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2)$, $t = 1$: trial = -2, diff = -2

Compute :

$$\text{A: } w_1^{t+1} = 2$$

$$\text{B: } w_1^{t+1} = w_1^t + [\text{diff}]s_t a_t = 0 + (-2) \cdot 1 \cdot 0 = 0$$

$$\text{C: } w_1^{t+1} = 1$$

$$\text{D: } w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2)$, $t = 1$: trial = -2, diff = -2 $\Rightarrow w_1^{t+1} = 0$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = 1$$

$$C: w_0^{t+1} = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2)$, $t = 1$: trial = -2, diff = -2 $\Rightarrow w_1^{t+1} = 0$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = 1$$

$$C: w_0^{t+1} = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2)$, $t = 1$: trial = -2, diff = -2 $\Rightarrow w_1^{t+1} = 0$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = 1$$

$$C: w_0^{t+1} = w_0^t + [\text{diff}](1 - a_t) = 0 + -2(1 - 1) = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 1$ $\mathbf{w} = (w_1, w_0) = (0, 0)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 1$ $\mathbf{w} = (w_1, w_0) = (0, 0)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2:$

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$:

Compute:

A: trial = -2

B: trial=0

C: trial = -1

D: trial = $2 + \max\{0, 0\} = 2$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$: trial = 2

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 2

C: diff = -1

D: diff = -2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$: trial = 2

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 2 - 0 = 2

C: diff = -1

D: diff = -2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$: trial = 2, diff = 2

Compute :

$$\text{A: } w_1^{t+1} = 2$$

$$\text{B: } w_1^{t+1} = 0$$

$$\text{C: } w_1^{t+1} = 1$$

$$\text{D: } w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$: trial = 2, diff = 2

Compute :

$$\text{A: } w_1^{t+1} = 0 + 2 \cdot 1 \cdot 1 = 2$$

$$\text{B: } w_1^{t+1} = 0$$

$$\text{C: } w_1^{t+1} = 1$$

$$\text{D: } w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 1$: trial = 2, diff = 2 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = 1$$

$$C: w_0^{t+1} = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 1$: trial = 2, diff = 2 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$\text{A: } w_0^{t+1} = 2$$

$$\text{B: } w_0^{t+1} = 1$$

$$\text{C: } w_0^{t+1} = 0$$

$$\text{D: } w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 2$: trial = 2, diff = 2 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = 1$$

$$C: w_0^{t+1} = 0 + 2(1 - 1) = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 2$ $\mathbf{w} = (w_1, w_0) = (2, 0)$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3:$

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0 + \max\{(2 \cdot (-1) \cdot 0 + 0(1 - 0)), (2(-1)1 + 0(1 - 1))\} = 0 + \max\{-2, 0\} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0)$, $t = 3$: trial = 0

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 2

C: diff = -1

D: diff = -2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 2$ $\mathbf{w} = (w_1, w_0) = (2, 0)$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0)$, $t = 3$: $\text{trial} = 0$

Compute $\text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$:

A: $\text{diff} = 0 - (2 \cdot 0 \cdot 0 + 0(1 - 0)) = 0$

B: $\text{diff} = 2$

C: $\text{diff} = -1$

D: $\text{diff} = -2$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0)$, $t = 3$: trial = 0, diff = 0

Since [diff]= 0:

\Rightarrow no change in (w_1, w_0)

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$:

Compute:

$$\text{A: trial} = -2$$

$$\text{B: trial} = 0$$

$$\text{C: trial} = -1 + 0 = -1$$

$$\text{D: trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4: \text{trial} = -1$

Compute $\text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$:

A: $\text{diff} = 0$

B: $\text{diff} = 2$

C: $\text{diff} = -1$

D: $\text{diff} = -2$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 3$ $\mathbf{w} = (w_1, w_0) = (2, 0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$: trial = -1

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 2

C: diff = $-1 - 0 = -1$

D: diff = -2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4: \text{trial} = -1, \text{diff} = -1$

Compute :

$$\text{A: } w_1^{t+1} = 2$$

$$\text{B: } w_1^{t+1} = 0$$

$$\text{C: } w_1^{t+1} = 1$$

$$\text{D: } w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$: trial = -1, diff = -1

Compute :

$$A: w_1^{t+1} = 2 + (-1) \cdot (-1) \cdot 0 = 2$$

$$B: w_1^{t+1} = 0$$

$$C: w_1^{t+1} = 1$$

$$D: w_1^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1), t = 4$: trial = -1, diff = -1 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$A: w_0^{t+1} = 2$$

$$B: w_0^{t+1} = -1$$

$$C: w_0^{t+1} = 0$$

$$D: w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1)$, $t = 4$: trial = -1, diff = -1 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$\text{A: } w_0^{t+1} = 2$$

$$\text{B: } w_0^{t+1} = -1$$

$$\text{C: } w_0^{t+1} = 0$$

$$\text{D: } w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = -1, a_t = 0, s_{t+1} = \text{exit}, r_{t+1} = -1)$, $t = 4$: trial = -1, diff = -1 $\Rightarrow w_1^{t+1} = 2$

Compute :

$$\text{A: } w_0^{t+1} = 2$$

$$\text{B: } w_0^{t+1} = 0 + (-1) \cdot (1 - 0) = -1$$

$$\text{C: } w_0^{t+1} = 0$$

$$\text{D: } w_0^{t+1} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 4$ $\mathbf{w} = (w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5:$

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 4$ $\mathbf{w} = (w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5:$

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5$:

Compute:

$$A: \text{trial} = -2$$

$$B: \text{trial} = 0$$

$$C: \text{trial} = -1$$

$$D: \text{trial} = 2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

► $\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

► $\text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$

$t = 4$ $\mathbf{w} = (w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5: \text{trial} = 2$

Compute $\text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$:

A: $\text{diff} = 0$

B: $\text{diff} = 2$

C: $\text{diff} = -1$

D: $\text{diff} = -2$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2), t = 5$: trial = 2

Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

$$\text{A: diff} = 2 - (2 \cdot 1 \cdot 1 + (-1)(1 - 1)) = 2 - 2 = 0$$

$$\text{B: diff} = 2 - 0 = 2$$

$$\text{C: diff} = -1$$

$$\text{D: diff} = -2$$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2)$, $t = 5$: trial = 2, diff = 0

Since [diff]= 0:

\Rightarrow no change in (w_1, w_0)

Final solution: $\mathbf{w} = (w_1, w_0) = (2, -1)$

Approximative Q-learning

Episode 1	Episode 2	Episode 3
(0, 1, 1, -2)	(0, 0, -1, 0)	(1, 1, exit, 2)
(1, 1, exit, 2)	(-1, 0, exit, -1)	

each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$

$$S = \{-1, 0, 1\}$$

$$A = \{0, 1\}$$

$$\gamma = 1, \alpha = 1$$

$$\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$$

Task: compute Q-function - from each tuple refine w_0, w_1

$$\blacktriangleright \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{diff})\nabla\hat{q}(s_t, a_t, \mathbf{w}); \text{diff} = \text{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

$$w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

$$\blacktriangleright \text{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

$$t = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

Transition $(s_t = 1, a_t = 1, s_{t+1} = \text{exit}, r_{t+1} = 2)$, $t = 5$: trial = 2, diff = 0

Since [diff]= 0:

\Rightarrow no change in (w_1, w_0)

Final solution: $\mathbf{w} = (w_1, w_0) = (2, -1)$