Learning by Approximation

J. Kostlivá, Z. Straka, P. Švarný

Today two examples:

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

How?:

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

How?:

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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: $\sum_{i}(\hat{f}(x_i,\mathbf{w})-f(x_i))$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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: $\sum_{i}(\hat{f}(x_i, \mathbf{w}) f(x_i))$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $E = -\infty$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $E = -\infty$

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: $\partial E = -\infty$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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: $\partial E = -\infty$

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**: $\partial E = -\infty$

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: w
- C: w₁
- D: $f(x_i)$

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: **w**

 $C: w_1$

D: $f(x_i)$

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: *×*

B: **w**

C: *W*₁

D: $f(x_i)$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

Evaluate $\frac{\partial L}{\partial w_0}$

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

B:
$$\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i + 1 - f(x_i))$$

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$ Evaluate $\frac{\partial E}{\partial w_0}$:

- A: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i f(x_i))$
- B: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i + 1 f(x_i))$
- C: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i + w_0 f(x_i))$
- D: $\frac{\partial E}{\partial w_0} = 2 \sum_i (x_i f(x_i))$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$ Evaluate $\frac{\partial E}{\partial w_0}$:

- A: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i f(x_i))$
- B: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i + 1 f(x_i))$
- C: $\frac{\partial E}{\partial w_0} = 2 \sum_i (w_1 x_i + w_0 f(x_i))$
- D: $\frac{\partial E}{\partial w_0} = 2 \sum_i (x_i f(x_i))$

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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

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

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

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

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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

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

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

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

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

We have:

- **v** given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- **>** approximation of function $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Solve linear equation system.

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

We have:

- **v** given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- **>** approximation of function $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Solve linear equation system.

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

We have:

- **v** given tuples $(x_i, f(x_i)) : (0, 2.1), (1, 3.6), (2, 4.9), (3, 6.6), \dots$
- ▶ approximation of function $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Solve linear equation system.

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Solve linear equation system.

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

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

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

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

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$
 $\frac{\partial E}{\partial w_1} = \sum_{i} (w_1 x_i + w_0 - f(x_i)) x_i = 0$

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

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 $\hat{f}(x, \mathbf{w}) = w_1x + w_0$

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$

Evaluate

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

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_i (w_1 x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$
 $\frac{\partial E}{\partial w_1} = \sum_i (w_1 x_i + w_0 - f(x_i))x_i = 0$

Evaluate:

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

 $-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$ $w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$

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

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

 $-2w_1 + 2.8 = 0 \rightarrow w_1 = 1.4$ $w_0 = 1/3(10.6 - 3w_1) = \frac{6.4}{3} \approx 2.133$

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

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

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

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

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

 $w_0 = 1/3(10.6 - 3w_1) = \frac{0.4}{3} \approx 2.133$

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

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

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

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

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

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

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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

We have:

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

Today:

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

We have:

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

Today:

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

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

 $S = \{-1, 0, 1\}$ $A = \{0, 1\}$ $\hat{q}(s, a, w) = asw_1 + (1 - a)w_0$

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

- Find 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 w that value of $(\text{trial}_t \hat{q}(s_t, a_t, w))^2$ will decrease
- ► How?

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, w) = asw_1 + (1 - a)w$

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

- Find 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 **w** that value of $(\text{trial}_t \hat{q}(s_t, a_t, \mathbf{w}))^2$ will decrease
- ► How

	Episode 1	Episo	de 1 Episo	de 2	Episode 3
(1 1 ': 0) (1 0 ': 1)	(0,1,1,-2)	0, 1, 1	(0,0,-1)	-1,0)	(1, 1, exit, 2)
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(1, 1, exit, 2)	1, 1, e	(-1,0,e)	(it, -1)	

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

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

 $A = \{0, 1\}$
 $a(s, a, w) = asw_0 + (1 - a)w_0$

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

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

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
$(1,1,exit,2) \mid (-1,0,exit,-1) \mid$		
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

- Find 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 w that value of $(trial_t \hat{q}(s_t, a_t, w))^2$ will decrease
- ► How

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
$(1,1,exit,2) \mid (-1,0,exit,-1) \mid$		
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 wo, we

- Find 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 w that value of $(trial_t \hat{q}(s_t, a_t, w))^2$ will decrease
- ► How

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
$(1,1,exit,2) \mid (-1,0,exit,-1) \mid$		
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

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

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
$(1,1,exit,2) \mid (-1,0,exit,-1) \mid$		
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

- ▶ Find **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?

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	
and field in the table is an intended of a second		

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

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	
and field in the table is an interval (a. a. a. iii)		

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

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

Define

- A: trial = $r_{t+1} + \gamma \hat{q}(s_{t+1}, a, \mathbf{w})$
- B: trial = $r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
- C: trial = $\gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
- D: trial = $r_{t+1} + \gamma \max_a \hat{q}(s_t, a, \mathbf{w})$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

Define:

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

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

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

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

Define:

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

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

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

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

(0,1,1,-2) $(0,0,-1,0)$ $(1,1,exit,1)$	3	E		1	de 1	pisod	E
	2)	0		2)	(1, -2)	0, 1, 1,	(0
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$		1	(-	2)	exit, 2)	1, ex	(1,

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

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

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

Define w₁ 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: $w_1^{t+1} = w_1^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

(0,1,1,-2) $(0,0,-1,0)$ $(1,1,exit)$	
	, 2)
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	

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

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

- $ightharpoonup \mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{trial} \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w})$
- lacksquare 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:
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))$$

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

(0,1,1,-2) $(0,0,-1,0)$ $(1,1,exit)$	
	, 2)
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	

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

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

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

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

 $A = \{0, 1\}$
 $\hat{a}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1,1,exit,2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

Define w_0 update:

A:
$$\mathbf{w}_0^{t+1} = \mathbf{w}_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))$$

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	
1 (11 1 1 1	11 1 1 /	

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

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

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

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

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

ightharpoonup 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: $\gamma=1,lpha=1$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

ightharpoonup 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: $\gamma = 1, \alpha = 1$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 (11 1 1 1	11 ' ' 1 (`

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

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

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

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

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

lacksquare 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: $\gamma = 1, \alpha = 1$

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

 $A = \{0, 1\}$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

Initialize 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1$, $\alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 (1 1 1 1 1 1	11 /	

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

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

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

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

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
() / / () - / - / - / - /	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t=0$$
 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:
$$trial = -2$$

$$B: trial = 0$$

$$C: trial = -i$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

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

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:
$$trial = 1$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

(0,1,1,-2) $(0,0,-1,0)$ $(1,1,ex)$	le 3
	it, 2)
(1, 1, exit, 2) (-1, 0, exit, -1)	

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

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

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

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

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

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

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

B: trial=0

C: trial = -1

D: trial = 1

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))(1 - \mathbf{a}_t)$$

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

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

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:
$$diff = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) \mathbf{s}_t \mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) (1 - \mathbf{a}_t)$$

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

$$t=0$$
 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \alpha = 1$
 $\hat{a}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 (1 1 1 1 1 1	11 /	`

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))(1 - \mathbf{a}_t)$$

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

$$t=0$$
 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 $\Rightarrow w_1^{t+1} = 0$

-
- A: $w_0^{t+1} = 2$
- B: $w_0^{t+1} = 1$
- C: $w_0^{t+1} = 0$
- D: $w_0^{t+1} = -2$

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

 $A = \{0, 1\}$
 $\gamma = 1, \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
() / / () - / - / - / - /	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) \mathbf{s}_t \mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) (1 - \mathbf{a}_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0,0)$

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

- A: trial = -2
- B: trial = 0
- C: trial = -1
- D: trial = 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 6 1 1 1 1	11 1 1 /	

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

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

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

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

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

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

$$t=1$$
 w = $(w_1, w_0) = (0,0)$

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

Compute

A: trial = -2

B: trial = 0

C: trial = -1

D: trial = 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))(1 - \mathbf{a}_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

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

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C: trial =
$$-1$$

D:
$$trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

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

Compute:

A:
$$trial = -2$$

C: trial =
$$-1$$

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))(1 - \mathbf{a}_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0,0)$

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\qquad \qquad \mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

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

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

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

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

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

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

D:
$$W_1^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 (1 1 1 1 1 1	11 /	`

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

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

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

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

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

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

$$t=1$$
 w = $(w_1, w_0) = (0,0)$

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

- A ++-1
- A: $w_0^{t+1} = 2$
- B: $w_0^{t+1} = 1$
- C: $w_0^{t+1} = 0$
- D: $w_0^{t+1} = -2$

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

 $A = \{0, 1\}$
 $\gamma = 1, \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0, 0)$

Transition
$$(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 1$$
: 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:
$$w_0^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t=1$$
 w = $(w_1, w_0) = (0,0)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 2$: 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:
$$w_0^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t=2$$
 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 (1 1 1 1 1 1	11 /	

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

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

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

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

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

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

$$t=2$$
 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:
$$trial = 0$$

D:
$$trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

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

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

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

$$t = 2$$
 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
$$= 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
(1.1 evit 2) (-1.0 evit -1)	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

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

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

$$t=2$$
 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:
$$trial=0 + max\{(2 \cdot (-1) \cdot 0 + 0(1-0)), (2(-1)1 + 0(1-1))\} = 0 + max\{-2, 0\} = 0$$

C: trial =
$$-1$$

$$D: trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t = 2$$
 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t$$

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

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

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

$$t=2$$
 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 =
$$0 - (2 \cdot 0 \cdot 0 + 0(1 - 0)) = 0$$

B: diff = 2

C: diff = -1

D: diff = -2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 t	able is an n-tuple (s_t , a	(s_t, s_{t+1}, r_{t+1})

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

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

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

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

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

$$t=2$$
 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)

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition ($s_t = -1$, $a_t = 0$, $s_{t+1} = exit$, $r_{t+1} = -1$), t = 4: Compute:

- A: trial = -2
- B: trial = 0
- C: trial = -1
- D: trial = 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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

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

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

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

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition
$$(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$$
:

Compute

A:
$$trial = -2$$

B:
$$trial = 0$$

$$C: trial = -i$$

D:
$$trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$:

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C: trial =
$$-1$$

D:
$$trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition ($s_t = -1$, $a_t = 0$, $s_{t+1} = exit$, $r_{t+1} = -1$), t = 4: Compute:

A:
$$trial = -2$$

B: trial=0

C: trial =
$$-1 + 0 = -1$$

D: trial = 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$

$$w_1 = w_1 + \alpha(\operatorname{trial} - q(s_t, a_t, \mathbf{w}))s_t a_t$$

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$: trial = -1 Compute diff $= trial - \hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

B: diff = 2

C: diff = -1

D: diff = -2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) \mathbf{s}_t \mathbf{a}_t$$

$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t)) (1 - \mathbf{a}_t)$$

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$: 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
(1.1 evit 2) (-1.0 evit -1)	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t = 3$$
 w = $(w_1, w_0) = (2, 0)$

Transition
$$(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$$
: trial = -1, diff = -1 Compute :

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

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

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

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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition
$$(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$$
: 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:
$$W_1^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t=3$$
 w = $(w_1, w_0) = (2,0)$

Transition $(s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1), t = 4$: trial = -1. diff $= -1 \Rightarrow w_t^{t+1} = 2$

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

 $A = \{0, 1\}$
 $\gamma = 1, \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
(1.1 evit 2) (-1.0 evit -1)	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t = 3$$
 w = $(w_1, w_0) = (2, 0)$

Transition (
$$s_t = -1, a_t = 0, s_{t+1} = exit, r_{t+1} = -1$$
), $t = 4$: 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:
$$w_0^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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)	
1 6 1 1 1 1	11 1 1 /	

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

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

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

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

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

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

$$t = 3$$
 w = $(w_1, w_0) = (2, 0)$

Transition (
$$s_t = -1$$
, $a_t = 0$, $s_{t+1} = exit$, $r_{t+1} = -1$), $t = 4$: trial $= -1$, diff $= -1 \Rightarrow w_1^{t+1} = 2$ Compute :

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

B:
$$w_0^{t+1} = 0 + (-1) \cdot (1-0) = -1$$

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

D:
$$W_0^{t+1} = -2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
() / / () - / - / - / - /	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 5$: Compute:

- A: trial = -2
- B: trial = 0
- C: trial = -1
- D: trial = 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

	Episode 1	Episode 2	Episode 3
$(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$	(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
() / / () - / - / - / - /	(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	

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

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

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

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

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

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 5$:

Compute

A:
$$trial = -2$$

B:
$$trial = 0$$

$$C: trial = -i$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 5$:

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C: trial =
$$-1$$

D:
$$trial = 2$$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

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

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition $(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 5$:

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial= 2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition
$$(s_t = 1, a_t = 1, s_{t+1} = exit, r_{t+1} = 2), t = 5$$
: 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

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff})\nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

$$w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_t a_t w_0^{t+1} = w_0^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition ($s_t = 1$, $a_t = 1$, $s_{t+1} = exit$, $r_{t+1} = 2$), t = 5: trial = 2 Compute diff = trial - $\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff =
$$2 - (2 \cdot 1 \cdot 1 + (-1)(1 - 1)) = 2 - 2 = 0$$

B: diff = 2 - 0 = 2

C: diff = -1

D: diff = -2

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

Episode 1	Episode 2	Episode 3
(0,1,1,-2)	(0,0,-1,0)	(1, 1, exit, 2)
(1, 1, exit, 2)	$(-1,0,\mathit{exit},-1)$	
each field in the table is an n-tunle (s. a. s		

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

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

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

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

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

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition (
$$s_t = 1$$
, $a_t = 1$, $s_{t+1} = exit$, $r_{t+1} = 2$), $t = 5$: trial = 2, diff = 0
Since [diff] = 0:
 \Rightarrow no change in (w_1 , w_0)

 \Rightarrow no change in (w_1, w_0)

Final solution: $\mathbf{w} = (w_1, w_0) = (2, -1)$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$

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 (st. at. st. 1, rt. 1)		

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

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

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

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

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

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

$$t = 4$$
 w = $(w_1, w_0) = (2, -1)$

Transition ($s_t = 1$, $a_t = 1$, $s_{t+1} = exit$, $r_{t+1} = 2$), t = 5: trial = 2, diff = 0 Since [diff] = 0: \Rightarrow no change in (w_1 , w_0) Final solution: $\mathbf{w} = (w_1, w_0) = (2, -1)$

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

 $A = \{0, 1\}$
 $\gamma = 1, \ \alpha = 1$
 $\hat{q}(s, a, \mathbf{w}) = asw_1 + (1 - a)w_0$