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

- 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

- 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

- 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

- 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? - minimize sum of squared errors.

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: $E = \sum_{i} (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$ 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
$$\hat{f}(x, \mathbf{w}) = w_1 x + w_0$$

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: $E = \sum_{i} (\hat{f}(x_i, \mathbf{w}) - f(x_i))^2$ 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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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

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 $\partial E = 0$

A: *x*

B: w

C: w;

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 $\partial E = 0$ 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 $\hat{f}(x, \mathbf{w}) = w_1 x + w_0$

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 $\partial E = 0$ 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

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:

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

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

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 $\frac{\partial E}{\partial w}$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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: $\frac{\partial E}{\partial w_2} = 2w_0 - 3.1$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Least square approximation

We have:

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

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

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

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

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

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

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

Least square approximation

We have:

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

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

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

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

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

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

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

Least square approximation

We have:

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

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

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

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

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

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

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

We have:

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

Today:

- we approximate Q-function
- $\hat{q}(s,a,\mathbf{w}) = asw_1 + (1-a)w_0$
- we will compute parameters w₀, w₁

We have:

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

Today:

- we approximate Q-function
- $\blacktriangleright \hat{q}(s,a,\mathbf{w}) = asw_1 + (1-a)w_0$
- 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, exit, -1)	
and field in the t	- la la la sua de deserval a la sua)

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 w₀, w₁

- 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 (c. a. c. r .)		

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

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

Task: compute Q-function - from each tuple refine w₀, w₁

- Find w that minimize $\sum_t (\text{trial}_t \hat{q}(s_t, a_t, \mathbf{w}))^2$
- How to do it online?
- ln every timestep t, modify w that value of $(\text{trial}_t \hat{q}(s_t, a_t, w))^2$ 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)	$(-1,0,\mathit{exit},-1)$	
each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$		

$$\begin{split} S &= \{-1, 0, 1\} \\ A &= \{0, 1\} \\ d(s, a, w) &= as w + (1 - a) w \end{split}$$

Task: compute Q-function - from each tuple refine w₀, w₁

- 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)	$(-1,0,\mathit{exit},-1)$	
each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$		

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

Fask: compute Q-function - from each tuple refine w₀, w₁

- 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)	
	11.1	``

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

SGD briefly:

- 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?

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

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

	ode 2 Episode 3
(0,1,1,-2) $(0,0,-)$	(1, 1, exit, 2)
(1, 1, exit, 2) $(-1, 0, exit, 2)$	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
$$w_0, w_1$$

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

A =*q(s*,

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

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

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

- Find w that minimize $\sum_t (trial_t \hat{q}(s_t, a_t, \mathbf{w}))^2$
- 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)	(-1, 0, exit, -1)	
each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$		

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

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 t	able is an e-tuele (c	

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
$$w_0, w_1$$

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

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

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

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

Define:

```
A: 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, w)

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

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

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

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

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

Define:

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

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

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

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

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

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

Define:

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

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

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

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

Define w_1 update:

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

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

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

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

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

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

Define w_1 update:

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

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

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

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

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

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

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

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

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

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

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

Define w_0 update:

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ \mathbf{v}_1^{t+1} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$

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

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

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

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

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

Define w₀ 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: $w_0^{t+1} = w_0^t + (\text{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$

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

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 e-tuple (c.)	

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

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

Let's compute $\mathbf{w}=(w_1,w_0)$ For simplicity: $\gamma=1, lpha=1$

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

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 e-tuple (c.)	

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

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

Let's compute $\mathbf{w} = (w_1, w_0)$ For simplicity $\sigma = 1, \sigma = 1$

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

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 e-tuple (c.)	

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

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

Let's compute $\mathbf{w} = (w_1, w_0)$ For simplicity: $\gamma = 1, \alpha = 1$

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

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

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

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, 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

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

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 6 1 1 1 1	11 1 1 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

A: trial = -2

t

- 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)	
1 6 1 1 2 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

A: trial = -2

t

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

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

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

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

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

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

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

A: trial = -2

t

- 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

B: trial=0

C: trial = -1

t

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition (
$$s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2$$
), $t = 1$: trial = -2
Compute diff = trial $-\hat{q}(s_t, a_t, \mathbf{w})$:
A: diff = 0

B: diff = 1

t

- $\textbf{C:} \ diff = \textbf{-1}$
- 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

Transition ($s_t = 0, a_t = 1, s_{t+1} = 1, r_{t+1} = -2$), t = 1: trial = -2 Compute diff = trial $-\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

t

- **B**: diff = 1
- $\textbf{C:} \ diff = -1$

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, 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

B: $w_1^{t+1} = 0$
C: $w_1^{t+1} = 1$
D: $w_1^{t+1} = -2$

t

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

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

t

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$
$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$
$$\mathsf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$
$$= 0 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

t

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

A: trial = -2

t

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) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{v}_1^{t+1} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

A: trial = -2

t

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) $(0, 0, -1,$	(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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
▶ $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 1 $\mathbf{w} = (w_1, w_0) = (0, 0)$

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{v}_1^{t+1} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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 = $2 + \max\{0, 0\} = 2$

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

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

(0, 1, 1, -2) $(0, 0, -1,$	(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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
▶ $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 1 $\mathbf{w} = (w_1, w_0) = (0, 0)$

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
- $\textbf{C:} \ diff = \textbf{-1}$
- D: diff = -2

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

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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
▶ $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 1 $\mathbf{w} = (w_1, w_0) = (0, 0)$

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
- $\textbf{C:} \ diff = -1$
- **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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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, 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

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

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 1 \quad \mathbf{w} = (w_1, w_0) = (0, 0)$$

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: Compute:

A: trial = -2

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

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

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

(0, 1, 1, -2) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{v}_1^{t+1} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition ($s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0$), t = 3: Compute:

A: trial = -2

t

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition ($s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0$), t = 3: Compute:

A: trial = -2

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: Compute:

A: trial = -2

t

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: trial = 0 Compute diff = trial $-\hat{q}(s_t, a_t, \mathbf{w})$:

A: diff = 0

- B: diff = 2
- C: diff = -1
- D: diff = -2

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

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

(0, 1, 1, -2) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 2 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: trial = 0 Compute diff = trial $-\hat{q}(s_t, a_t, \mathbf{w})$: A: diff = $0 - (2 \cdot 0 \cdot 0 + 0(1 - 0)) = 0$

B: diff = 2

t

 $\textbf{C:} \ diff = -1$

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) $(-1, 0, exit, -1)$	(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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))s_ta_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
▶ $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 2 $\mathbf{w} = (w_1, w_0) = (2, 0)$
Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: $\operatorname{trial} = 0, \operatorname{diff} = 0$

Fransition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: trial = 0, diff = Since [diff] = 0: \Rightarrow no change in (w_1, w_0)

$$egin{aligned} S &= \{-1,0,1\} \ A &= \{0,1\} \ \gamma &= 1, \ lpha &= 1 \ \hat{q}(s,a,\mathbf{w}) &= asw_1 + (1-a)w_0 \end{aligned}$$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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

A: trial = -2

t

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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, 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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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

- A: trial = -2
- **B**: trial=0

t

C: trial = -1 + 0 = -1

D: trial = 2

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$
$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) (1 - a_t)$$
$$\mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$
$$= 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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

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

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

(0, 1, 1, -2) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$
$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) (1 - a_t)$$
$$\mathbf{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$$
$$= 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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
- $\textbf{C:} \ diff = -1-0 = -1$
- **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
(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

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

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

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

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

$$= 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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

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

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

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

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

$$= 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
► $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 3 $\mathbf{w} = (w_1, w_0) = (2, 0)$

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

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

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

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

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

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

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

$$= 3 \quad \mathbf{w} = (w_1, w_0) = (2, 0)$$

$$egin{aligned} S &= \{-1,0,1\} \ A &= \{0,1\} \ \gamma &= 1, \ lpha &= 1 \ \hat{q}(s,a,\mathbf{w}) &= asw_1 + (1-a)w_0 \end{aligned}$$

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

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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
► $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 3 $\mathbf{w} = (w_1, w_0) = (2, 0)$

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

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

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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
▶ $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 4 $\mathbf{w} = (w_1, w_0) = (2, -1)$

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

A: trial = -2

t

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) $(0, 0, -1,$	(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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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

A: trial = -2

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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

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

 $w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$
 $w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$
► $\operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w})$
= 4 $\mathbf{w} = (w_1, w_0) = (2, -1)$

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

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

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

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t) \\ \mathbf{b} \quad \operatorname{trial} = r_{t+1} + \gamma \max_a \hat{q}(s_{t+1}, a, \mathbf{w}) \\ = 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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

 $\textbf{C:} \ diff = -1$

t

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

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

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

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

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

$$= 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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: $w = (w_1, w_0) = (2, -1)$

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

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

t

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

$$\mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})$$
$$w_1^{t+1} = w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t$$
$$w_0^{t+1} = w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t))(1 - a_t)$$
$$\mathbf{trial} = r_{t+1} + \gamma \max_{a} \hat{q}(s_{t+1}, a, \mathbf{w})$$
$$= 4 \quad \mathbf{w} = (w_1, w_0) = (2, -1)$$

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

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

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

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

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