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

Task: $\mathsf{determine}/\mathsf{compute}$ parameters w_0, w_1 with lowest errorr

How?:

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

We have:

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

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

How?:

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

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

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

How?:

A: minimize difference in coordinates

B: maximize error

C: minimize sum of squared errors

D: maximize difference in coordinates

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

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

How?:

A: minimize difference in coordinates

B: maximize error

C: minimize sum of squared errors

D: maximize difference in coordinates

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

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

How? - minimize sum of squared errors.

Define:

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

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

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

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

How? - minimize sum of squared errors. Define:

- A: $\sum_{i} (f(x_i) x_i)^2$
- B: $\sum_i (\hat{f}(x_i, w) f(x_i))^2$
- C: $\sum_i (x_i f(x_i))^2$
- D: $\sum_{i}(\hat{f}(x_i, w) f(x_i))$

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

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

How? - minimize sum of squared errors.

Define:

- A: $\sum_{i} (f(x_i) x_i)^2$
- B: $\sum_i (\hat{f}(x_i, w) f(x_i))^2$
- C: $\sum_{i} (x_i f(x_i))^2$
- D: $\sum_{i}(\hat{f}(x_i, w) f(x_i))$

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

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

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

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

We have:

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

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\partial E = 0$

Derive by

- A: x
- B: w
- C: w₁
- D: $f(x_i)$

We have:

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

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

We have:

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

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_{i} (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$ 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 $\hat{f}(x, w) = w_1x + w_0$

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

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

We have:

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

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

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

We have:

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

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

Minimize sum of squared errors: $E = \sum_i (\hat{f}(x_i, w) - f(x_i))^2$ Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$

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

Solve linear equation system

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

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) x_i = 0$

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) x_i = 0$

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$
 $\frac{\partial E}{\partial w_1} = 2 \sum_{i} (w_1 x_i + w_0 - f(x_i)) x_i = 0$

Solve linear equation system.

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

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Evaluate

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

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

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

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

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Evaluate:

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

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

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

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

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_{i} (w_1 x_i + w_0 - f(x_i)) = 0$
 $\frac{\partial E}{\partial w_1} = \sum_{i} (w_1 x_i + w_0 - f(x_i)) x_i = 0$

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:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_{i} (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

Evaluate

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

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

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

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

We have:

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

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_i (w_1 x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$
 $\frac{\partial E}{\partial w_0} = \sum_i (w_1 x_i + w_0 - f(x_i)) = 3w_1 + 3w_0 - 10.6 = 0$
 $\frac{\partial E}{\partial w_1} = \sum_i (w_1 x_i + w_0 - f(x_i))x_i = 0$

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

Least square approximation

We have:

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

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

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

Find minimum of E by derivation $\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \sum_i (w_1 x_i + w_0 - f(x_i))^2 = 0$

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

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

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

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

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

We have:

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

Today:

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

We have:

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

Today:

- we approximate Q-function
- $\hat{q}(s, a, w) = asw_1 + (1 a)w_0$
- 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) | $(-1,0,\mathit{exit},-1)$ | |

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

 $S = \{-1, 0, 1\}$ $A = \{0, 1\}$ $\hat{q}(s, a, w) = asw_1 + (1 - a)w_1$

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

- Find w that minimize $\sum_{t} (\text{trial}_{t} \hat{q}(s_{t}, a_{t}, w))^{2}$
- 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) | $(-1,0,\mathit{exit},-1)$ | |
| 1 (11 1 1 1 | 11 ' ' 1 (| |

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

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

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

- Find w that minimize $\sum_{t} (\text{trial}_t \hat{q}(s_t, a_t, w))^2$
- 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

| (0.1.1.0) $(0.0.1.0)$ $(1.1.1.0)$ | |
|---|----------|
| $\mid (0,1,1,-2) \mid (0,0,-1,0) \mid (1,1,\epsilon)$ | exit, 2) |
| $(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$ | |

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

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

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

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

- Find w that minimize $\sum_t (\text{trial}_t \hat{q}(s_t, a_t, w))^2$
- 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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

- ightharpoonup Find w that minimize $\sum_t (\operatorname{trial}_t \hat{q}(s_t, a_t, w))^2$
- ► How to do it online?
- ln every timestep t, modify w that value of $(\operatorname{trial}_t \hat{q}(s_t, a_t, w))^2$ 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) | $(-1,0,\mathit{exit},-1)$ | |
| 1 (11 1 1 1 | 11 ' ' 1 (| |

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

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

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

Task: compute Q-function - from each tuple refine wo, we

- Find w that minimize $\sum_t (\text{trial}_t \hat{q}(s_t, a_t, w))^2$
- ► 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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

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

- Find w that minimize $\sum_t (\text{trial}_t \hat{q}(s_t, a_t, w))^2$
- ► 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 $(s_t, a_t, s_{t+1}, r_{t+1})$ | | |

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

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

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

- Find w that minimize $\sum_t (\text{trial}_t \hat{q}(s_t, a_t, w))^2$
- ► 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 $(s_t, a_t, s_{t+1}, r_{t+1})$

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

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

How?:

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

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

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

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

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

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

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

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

How?:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Define w_1 update

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

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

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

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

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

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

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

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

Define w_1 update:

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

B:
$$\mathbf{w}_1^{t+1} = \mathbf{w}_1^t + \alpha(\text{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))$$

C:
$$\mathbf{w}_1^{t+1} = \mathbf{w}_1^t + \alpha(\text{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

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

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

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

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

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

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

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

Define w_1 update:

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

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

C:
$$\mathbf{w}_1^{t+1} = \mathbf{w}_1^t + \alpha(\text{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w}^t))\mathbf{s}_t\mathbf{a}_t$$

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

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

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

| 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

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

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

 $\qquad \qquad \mathsf{trial} = \mathit{r}_{t+1} + \gamma \, \mathsf{max}_{\mathsf{a}} \, \hat{q}(\mathit{s}_{t+1}, \mathit{a}, \mathsf{w})$

Define w_0 update

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

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

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

| 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) | |
| anala fialal in the t | - | |

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

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

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

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

Define w_0 update:

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

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

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

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

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

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

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

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

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

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

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

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

Define w_0 update:

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

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

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

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

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

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

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

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

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

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

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

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

| Episode 1 | Episode 2 | Episode 3 |
|-----------------|----------------|-----------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| (1, 1, exit, 2) | (-1,0,exit,-1) | |
| 1 (11 1 1 1 | 11. | |

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

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

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

Let's compute
$$w = (w_1, w_0)$$

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

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

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

| Episode 1 | Episode 2 | Episode 3 |
|-----------------|----------------|-----------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| (1, 1, exit, 2) | (-1,0,exit,-1) | |
| 1 (11 1 1 1 | 11 ' ' 1 (| ` |

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

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

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

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

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

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

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

| 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

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} + \alpha (\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w})) \nabla \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} &= w_1^t + \alpha (\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} &= w_0^t + \alpha (\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) (1 - a_t) \end{aligned}$$

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

Initialize w:

A:
$$w = (w_1, w_0) = (1, 1)$$

B:
$$w = (w_1, w_0) = (0, 1)$$

C:
$$w = (w_1, w_0) = (0, 0)$$

D: arbitrarily

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

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

| 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

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

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

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

Initialize w:

A:
$$w = (w_1, w_0) = (1, 1)$$

B:
$$w = (w_1, w_0) = (0, 1)$$

C:
$$w = (w_1, w_0) = (0, 0)$$

D: arbitrarily (we choose
$$w = (w_1, w_0) = (0, 0)$$
)

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

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

| 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

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

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

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

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

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

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

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

| 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

$$\begin{aligned} & \mathsf{w} \leftarrow \mathsf{w} + \alpha(\mathrm{trial} - \hat{q}(s_t, a_t, \mathsf{w})) \nabla \hat{q}(s_t, a_t, \mathsf{w}) \\ & w_1^{t+1} = w_1^t + \alpha(\mathrm{trial} - \hat{q}(s_t, a_t, \mathsf{w}^t)) s_t a_t \\ & w_0^{t+1} = w_0^t + \alpha(\mathrm{trial} - \hat{q}(s_t, a_t, \mathsf{w}^t)) (1 - a_t) \end{aligned}$$

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

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

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

Compute

A:
$$trial = -2$$

B:
$$trial = 0$$

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

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

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

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

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

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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C:
$$trial = -1$$

D:
$$trial = 1$$

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

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

| (0,1,1,-2) $(0,0,-1,0)$ $(1,1,e)$ | de 3 |
|---|---------|
| | xit, 2) |
| $(1, 1, exit, 2) \mid (-1, 0, exit, -1) \mid$ | |

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

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

$$ightharpoonup \mathbf{w} \leftarrow \mathbf{w} + \alpha(\text{trial} - \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})) \nabla \hat{q}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{w})$$

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

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

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

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

Compute:

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

B: trial=0

C: trial = -1

D: trial = 1

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

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

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

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

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

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

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

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

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

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

A:
$$diff = 0$$

$$B: diff = 1$$

C:
$$diff = -1$$

D: diff =
$$-2$$

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

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

| 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

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

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

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

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

A: diff = 0

B: diff = 1

C: diff = -1

D: diff = -2 - 0 = -2

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

Compute

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

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 = -:

D: trial = 2

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

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

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

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

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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C:
$$trial = -1$$

D:
$$trial = 2$$

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

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

| 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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

C: trial =
$$-1$$

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

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

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

| Episode 1 | Episode 2 | Episode 3 |
|-----------------|---------------------------|-----------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| (1, 1, exit, 2) | $(-1,0,\mathit{exit},-1)$ | |
| 1 6 11 1 | | |

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

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

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

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

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

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

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

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

A:
$$diff = 0$$

B:
$$diff = 2$$

C:
$$diff = -1$$

D:
$$diff = -2$$

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

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

| 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

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

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

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

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

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

A: diff = 0

B: diff =
$$2 - 0 = 2$$

C:
$$diff = -1$$

D:
$$diff = -2$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

Compute

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| Episode 1 | Episode 2 | Episode 3 |
|-----------------|---------------------------|-----------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| (1, 1, exit, 2) | $(-1,0,\mathit{exit},-1)$ | |
| 1 6 11 1 | | |

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

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

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

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

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

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

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

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

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

C:
$$w_0^{t+1} = 0 + 2(1-1) = 0$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

$$t=2 \text{ w} = (w_1, w_0) = (2, 0)$$

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 = -:

D: trial = 2

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

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

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

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

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

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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C:
$$trial = -1$$

D:
$$trial = 2$$

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

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

| 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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

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

C: trial =
$$-1$$

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

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

| 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

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

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

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

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

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

A:
$$diff = 0$$

B:
$$diff = 2$$

C:
$$diff = -1$$

D:
$$diff = -2$$

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

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

| 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

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

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

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

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

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

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

B: diff = 2

C: diff = -1

D: diff = -2

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

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

| Episode 1 | Episode 2 | Episode 3 |
|--|------------|-----------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| $(1,1,exit,2) \mid (-1,0,exit,-1) \mid$ | | |
| each field in the table is an n-tuple $(s_t, a_t, s_{t+1}, r_{t+1})$ | | |

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

$$\begin{aligned} \mathbf{w} \leftarrow \mathbf{w} + \alpha(\operatorname{diff}) \nabla \hat{q}(s_t, a_t, \mathbf{w}); & \operatorname{diff} = \operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}) \\ w_1^{t+1} &= w_1^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) s_t a_t \\ w_0^{t+1} &= w_0^t + \alpha(\operatorname{trial} - \hat{q}(s_t, a_t, \mathbf{w}^t)) (1 - a_t) \end{aligned}$$

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

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

Transition $(s_t = 0, a_t = 0, s_{t+1} = -1, r_{t+1} = 0), t = 3$: trial = 0, diff = 0 Since [diff]= 0:

 \Rightarrow no change in (w_1, w_0)

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

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

| 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

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

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

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

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

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

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

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

| 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

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

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

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

$$\qquad \qquad \mathsf{trial} = \mathit{r}_{t+1} + \gamma \, \mathsf{max}_{\mathit{a}} \, \hat{q}(\mathit{s}_{t+1}, \mathit{a}, \mathsf{w})$$

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

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

Compute

A:
$$trial = -2$$

B:
$$trial = 0$$

C:
$$trial = -1$$

D:
$$trial = 2$$

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

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

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

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

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

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

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

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

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

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

Compute:

A:
$$trial = -2$$

B:
$$trial = 0$$

C: trial =
$$-1$$

D:
$$trial = 2$$

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

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

| 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

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

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

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

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

A: trial = -2

B: trial=0

C: trial = -1 + 0 = -1

D: trial = 2

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

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

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

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

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

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

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

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

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

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

A:
$$diff = 0$$

B:
$$diff = 2$$

C:
$$diff = -1$$

D:
$$diff = -2$$

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

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

| 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

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

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

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

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

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

A: diff = 0

B: diff = 2

C: diff = -1 - 0 = -1

D: diff = -2

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

| Episode 1 | Episode 2 | Episode 3 |
|-----------------|---------------------------|---------------------------------------|
| (0,1,1,-2) | (0,0,-1,0) | (1, 1, exit, 2) |
| (1, 1, exit, 2) | $(-1,0,\mathit{exit},-1)$ | |
| 1 6 111 | | · · · · · · · · · · · · · · · · · · · |

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

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

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

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

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

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

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

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

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

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

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

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

| 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) | |
| | | L . |

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

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

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

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

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

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

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

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

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

| 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

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

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

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

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

Compute

A:
$$trial = -2$$

B:
$$trial = 0$$

D:
$$trial = 2$$

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

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

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

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

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

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

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

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

$$v_0 = v_0 + \alpha(v_1, u_1, v_2)$$

$$v_0 = v_0 + \alpha(v_1, u_1, v_2)$$

$$\qquad \qquad \mathsf{trial} = \mathit{r}_{t+1} + \gamma \, \mathsf{max}_{\mathsf{a}} \, \hat{q}(\mathit{s}_{t+1}, \mathit{a}, \mathsf{w})$$

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

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

Compute:

A:
$$trial = -2$$

$$B: trial = 0$$

C: trial =
$$-1$$

D:
$$trial = 2$$

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

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

| 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

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

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

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

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

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

Compute:

A: trial = -2

B: trial = 0

C: trial = -1

D: trial= 2

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

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

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

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

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

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

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

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

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

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

A:
$$diff = 0$$

B:
$$diff = 2$$

C:
$$diff = -1$$

D:
$$diff = -2$$

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

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

| 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

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

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

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

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

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

B: diff = 2 - 0 = 2

C: diff = -1

D: diff = -2

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

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

| 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

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

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

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

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

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

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

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

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

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

| | Episode 1 | Episode 2 | Episode 3 |
|---|--|-------------------|-----------------|
| ĺ | (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 $(s_t, a_t, s_{t+1}, r_{t+1})$ | | |

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
$$w_0, w_1$$

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

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

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

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

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

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

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

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