

1. Gradient descent is an algorithm for finding values of parameters  $w$  and  $b$  that minimize the cost function  $J$ .

repeat until convergence {

$$w = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w, b)$$

When  $\frac{\partial J(w, b)}{\partial w}$  is a negative number (less than zero), what happens to  $w$  after one update step?

- ☐  $w$  decreases
- ☒  $w$  increases.
- ☐ It is not possible to tell if  $w$  will increase or decrease.
- ☐  $w$  stays the same

☒ Correct

The learning rate is always a positive number, so if you take  $w$  minus a negative number, you end up with a new value for  $w$  that is larger (more positive).

2. For linear regression, what is the update step for parameter  $b$ ?

- ☒  $b = b - \alpha \frac{1}{m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})$
- ☐  $b = b - \alpha \frac{1}{m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})x^{(i)}$

☒ Correct

The update step is  $b = b - \alpha \frac{\partial J(w, b)}{\partial b}$  where  $\frac{\partial J(w, b)}{\partial b}$  can be computed with this expression:

$$\sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})$$