Practice quiz: Train the model with gradient descent

Total points 2

1. 1 point

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

2. 1 point

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

- ○ $b = b - \alpha \frac{1}{m} \sum\limits_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)}) x^{(i)}$
- ◉ $b = b - \alpha \frac{1}{m} \sum\limits_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)})$

Practice quiz: Supervised vs unsupervised learning

Total points 2

1. Which are the two common types of supervised learning? (Choose two) 1 point

- [x] Classification
- [ ] Clustering
- [x] Regression

2. 1 point

Which of these is a type of unsupervised learning?

- ○ Regression
- ○ Classification
- ◉ Clustering

Practice quiz: Regression

Total points 2

1. 1 point

For linear regression, the model is $f_{w,b}(x) = wx + b$.

Which of the following are the inputs, or features, that are fed into the model and with which the model is expected to make a prediction?

- (selected) x
- m
- (x, y)
- w and b.

2. For linear regression, if you find parameters w and b so that $J(w, b)$ is very close to zero, what can you conclude? 1 point

- (selected) The selected values of the parameters w and b cause the algorithm to fit the training set really well.
- The selected values of the parameters w and b cause the algorithm to fit the training set really poorly.
- This is never possible -- there must be a bug in the code.

Practice quiz: Multiple linear regression

Total points 4

1. In the training set below, what is $x_4^{(3)}$? Please type in the number below (this is an integer such as 123, no decimal points).

1 point

Size in feet^2	Number of bedrooms	Number of floors	Age of home in years	Price ($) in $1000's
X_1	X_2	X_3	X_4	
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

30

2.

1 point

Which of the following are potential benefits of vectorization? Please choose the best option.

- ( ) It can make your code shorter
- (•) All of the above
- ( ) It makes your code run faster
- ( ) It allows your code to run more easily on parallel compute hardware

3. True/False? To make gradient descent converge about twice as fast, a technique that almost always works is to double the learning rate $alpha$.

1 point

- ( ) True
- (•) False

4.

1 point

True/False? With polynomial regression, the predicted values f_w,b(x) does not necessarily have to be a straight line (or linear) function of the input feature x.

- (•) True
- ( ) False

Practice quiz: Gradient descent in practice

Total points 4

1. 

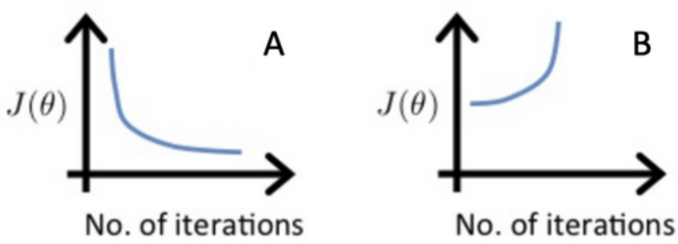
1 point

Which of the following is a valid step used during feature scaling?

- [x] Subtract the mean (average) from each value and then divide by the (max - min).
- [ ] Add the mean (average) from each value and and then divide by the (max - min).

2. Suppose a friend ran gradient descent three separate times with three choices of the learning rate α and plotted the learning curves for each (cost J for each iteration).

1 point



For which case, A or B, was the learning rate α likely too large?

- [ ] Both Cases A and B
- [x] case B only
- [ ] Neither Case A nor B
- [ ] case A only

3. Of the circumstances below, for which one is feature scaling particularly helpful?

1 point

- [x] Feature scaling is helpful when one feature is much larger (or smaller) than another feature.
- [ ] Feature scaling is helpful when all the features in the original data (before scaling is applied) range from 0 to 1.

4.

1 point

You are helping a grocery store predict its revenue, and have data on its items sold per week, and price per item. What could be a useful engineered feature?

- [x] For each product, calculate the number of items sold times price per item.
- [ ] For each product, calculate the number of items sold divided by the price per item.

1.

Gradient descent for logistic regression

1 / 1 point

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) \right]$$

} simultaneous updates

$$f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = \frac{1}{1 + e^{(-\vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b)}}$$

Which is the correct update step for

- [x] The update steps look like the update steps for linear regression, but the definition of $f_{\vec{w},b}(\mathbf{x}^{(i)})$ is different.
- [ ] The update steps are identical to the update steps for linear regression.

Correct
For logistic regression, $f_{\vec{w},b}(\mathbf{x}^{(i)})$ is the sigmoid function instead of a straight line.

Practice quiz: Cost function for logistic regression

Latest Submission Grade 100%

1.

1 / 1 point

$$\overbrace{J(\vec{\mathrm{w}},b)}^{?} = \frac{1}{m}\sum_{i=1}^{m} \underbrace{L\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right), y^{(i)}\right)}_{?}$$

In this lecture series, "cost" and "loss" have distinct meanings. Which one applies to a single training example?

- [x] Loss

Correct
In these lectures, loss is calculated on a single training example. It is worth noting that this definition is not universal. Other lecture series may have a different definition.

- [ ] Cost
- [ ] Both Loss and Cost
- [ ] Neither Loss nor Cost

2.

1 / 1 point

Simplified loss function

$$L\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$\boxed{L\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right), y^{(i)}\right) = -\,y^{(i)}\log\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right)}$$

For the simplified loss function, if the label $y^{(i)} = 0$, then what does this expression simplify to?

- ( ) $\log(f_{\vec{w},b}(\mathbf{x}^{(i)})$
- ( ) $\log(1 - f_{\vec{\mathbf{w}},b}(\mathbf{x}^{(i)})) + log(1 - f_{\vec{\mathbf{w}},b}(\mathbf{x}^{(i)}))$
- ( ) $-\log(1 - f_{\vec{\mathbf{w}},b}(\mathbf{x}^{(i)})) - log(1 - f_{\vec{\mathbf{w}},b}(\mathbf{x}^{(i)}))$
- (•) $-\log(1 - f_{\vec{\mathbf{w}},b}(\mathbf{x}^{(i)}))$

Correct
When $y^{(i)} = 0$, the first term reduces to zero.