Supervised Machine Learning: Regression and Classification

Week 1 to Week 3

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

Answer: Regression; Classification.

2 Which of these is a type of unsupervised learning?

Answer: Clustering.

3 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?

Answer: x.

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

Answer: The selected values of the parameters w and b cause the algorithm to fit the training set really well.

5 Gradient descent is an algorithm for finding values of parameters w and b that minimize the cost function J. When $\partial J(w,b)/\partial w$ is a negative number less than zero, what happens to w after one update step?

Answer: w increases.

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

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

7 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.

Answer: 30.

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

Answer: All of the above: it makes your code run faster; it can make your code shorter; it allows your code to run more easily on parallel compute hardware.

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

Answer: False.

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

Answer: Subtract the mean (average) from each value and then divide by the (max - min).

11 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).

Answer: Case B only.

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

Answer: Feature scaling is helpful when one feature is much larger (or smaller) than another feature.

13 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?

Answer: For each product, calculate the number of items sold times price per item.

14 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.

Answer: True.

15 Which is an example of a classification task?

Answer: Based on the size of each tumor, determine if each tumor is malignant (cancerous) or not.

16 Recall the sigmoid function is $g(z) = 1/(1+^{-z})$. If z is a large positive number, then

Answer: g(z) is near one.

17 A cat photo classification model predicts 1 if it's a cat, and 0 if it's not a cat. For a particular photograph, the logistic regression model outputs g(z) (a number between 0 and 1). Which of these would be a reasonable criteria to decide whether to predict if it's a cat?

Answer: Predict it is a cat if $g(z) \geq 0.5$.

18 True/False? No matter what features you use (including if you use polynomial features), the decision boundary learned by logistic regression will be a linear decision boundary.

Answer: False.