:::DATA::: QUIZ2

Given the task of learning to predict the output of a function given a particular input, you are given the following training examples of inputs for the function together with the correct output (x, F(x)):

(-7, -13),  (-3, -5),  (-1, -1),  (5, 11),  (7, 15)

We decide to model this as a linear function: F(x) = w0 + w1\*x

1) Initially choosing (w0, w1) = (0, 0)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

2) Initially choosing (w0, w1) = (0, 1)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

3) Initially choosing (w0, w1) = (1, 0)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

4) Initially choosing (w0, w1) = (1, 1)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

5) Initially choosing to try for (w0, w1) the four different models: (0,0), (1,0), (0, 1), (1, 1)

Which of these models is best according to Squared Loss? (w0,w1)??? WHAT WEIGHTS??

6) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 0.01? (Express to 2 decimal places)

7) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 0.01? (Express to 2 decimal places)

8) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 0.1? (Express to 2 decimal places)

9) Given the task of learning to predict the output of a function given a particular input, you are given the following training examples of inputs for the function together with the correct output (x, F(x)):

(-7, -13),  (-3, -5),  (-1, -1),  (5, 11),  (7, 15)

We decide to model this as a linear function: F(x) = w0 + w1\*x

Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 0.1? (Express to 2 decimal places)

10) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, which of the following learning rates is best?

Logistic:

Given the task of learning to determine between positive and negative examples of a class where each example has a single feature x.  Given these negative examples:

(-7), (-3), (-1)

and these positive examples:

(5), (7)

We decide to use Logistic Regression where we model positive as 1 and negative as 0 and guess positive using the model : Logistic(w0 + w1\*x)

11) Initially choosing (w0, w1) = (0, 0)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

12) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 1.0? (Express to 2 decimal places)

13) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 1.0? (Express to 2 decimal places)

14) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, using the weights after one update of the Gradient Descent algorithm, what would the model predict for example (-7): pos or neg??

15) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, using the weights after one update of the Gradient Descent algorithm, what would the model predict for example (7): pos or neg??

1. Between 540.5 and 541.5
2. Between 139.5 and 140.5
3. Between 531.5 and 532.5
4. Between 132.5 and 133.5
5. (1,1)
6. Between 0.004 and 0.024
7. Between 0.484 and 0.584
8. Between 0.04 and 0.24
9. Between 4.84 and 5.84
10. 0.01
11. Between 1.2 and 1.3
12. Between -0.15 and -0.05
13. Between 1.8 and 2.8
14. Pos
15. neg