Started on	Friday, February 24, 2023, 9:52 PM
State	Finished
Completed on	Friday, February 24, 2023, 10:07 PM
Time taken	14 mins 57 secs
Points	9.00/10.00
Grade	**90.00** out of 100.00

Question **1**

Correct

1.00 points out of 1.00

In batch gradient descent

Select one:

○ features, multiplied by their corresponding weights, are fed into the activation function in batches, so that the output of the activation function always reflects only a subset of the feature values for a particular sample.

○ several distinct hyperparameter settings can be tried at once, in a single batch job.

○ the gradient of the activation function is adjusted after each error is computed.

○ the threshold is lowered until the gradient of the activation function reaches zero.

◉ the errors for all training examples are calculated before the weights are updated. ✔

Your answer is correct.

The correct answer is: the errors for all training examples are calculated before the weights are updated.

Question **2**

Correct

1.00 points out of 1.00

While the perceptron has a learning rate hyperparameter, the Adaline model does not.

Select one:

○ True

◉ False ✔

The correct answer is 'False'.

With the Adaline model, higher learning rates always produce more accurate models, although they also make it necessary to run the learning algorithm for more iterations.

Select one:

○ True

◉ False ✔

The correct answer is 'False'.

In contrast to the Perceptron model, in the Adaline model

Select one:

○ the threshold is applied to each feature individually, and the features are considered in random order.

○ the average feature value is used as the threshold.

○ the weight vector is randomly permuted before each update.

○ the weights are initialized to 1.

◉ the activation function output, rather than the true class label, is used to compute Δw. ✔

Your answer is correct.

The correct answer is: the activation function output, rather than the true class label, is used to compute Δw.

If you are trying to train a classifier using all of the data in an enormous data set, then the large size of the data set would be a good reason to consider using

Select one:

○ batch gradient descent.

○ a learning rate of zero.

○ a perceptron model.

○ unsupervised learning.

◉ stochastic gradient descent.✔

○ a weight shuffling algorithm.

○ corpuscular reverberation.

Your answer is correct.

The correct answer is: stochastic gradient descent.

Minimizing the sum of squared errors objective function

Select one:

◉ might not minimize the number of examples misclassified.✔

○ is impossible when the data is linearly separable.

○ tends to maximize the standard deviation of the features.

○ will raise a "division by zero" exception.

○ prevents saturation of the activation function.

Your answer is correct.

The correct answer is: might not minimize the number of examples misclassified.

Question **7**

Correct

1.00 points out of 1.00

After feature standardization

Select one:

○ all features will be correlated with one another.

◉ the values of each feature will have a mean of zero and a standard deviation of one. ✔

○ changes to the random seed will have no effect.

○ no two features will share values.

○ at least one feature will have a constant value of zero.

Your answer is correct.

The correct answer is: the values of each feature will have a mean of zero and a standard deviation of one.

Question **8**

Incorrect

0.00 points out of 1.00

Which of the following computes a function that is equivalent to the activation function used in the Adaline model?

Select one:

○ import math
def activation(y, z):
    return 0.5 * math.sqrt(abs(y**2 - z**2))

◉ import math   ✘
def activation(y, z):
    return math.sqrt(y**2 - z**2)

○ def activation(x):
    return x * x

○ import math
def activation(y, z):
    return math.sqrt(abs(y**2 - z**2))

○ def activation(x):
    return x

Your answer is incorrect.

The correct answer is: def activation(x):
    return x

Question **9**

Correct

1.00 points out of 1.00

In the Adaline model with stochastic gradient descent,

Select one:

- ⦿ all weights may be adjusted after each training example is processed. ✔
- ○ the threshold is randomly decreased after each step of training.
- ○ the weights are randomized after each batch of training examples.
- ○ the weights are decreased until convergence.

Your answer is correct.

The correct answer is: all weights may be adjusted after each training example is processed.

Question **10**

Correct

1.00 points out of 1.00

When using stochastic gradient descent with the Adaline model,

Select one:

- ○ a step function is substituted for the linear activation function at random times, with increasing frequency over the learning process.
- ○ the threshold is driven by a white noise generator.
- ⦿ the training examples are processed in random order. ✔
- ○ the weights are randomized each time the error increases.
- ○ the training examples are processed from highest to lowest sum of feature values.

Your answer is correct.

The correct answer is: the training examples are processed in random order.

Jump to...