CS344 Exercise 5

Task 1: Advanced neural networks

For binary classification, a neural network with no hidden layers is equivalent to a logistic regression model.

TRUE FALSE

While neural networks may be used to predict whether data points are labeled 0 or 1, they cannot and should not be used to make predictions for data points whose labels take on a large range of decimal values.

TRUE FALSE

Suppose you have a neural network where, for one of the hidden layers, all units in the layer have identical values for their weight parameters as the other units in the layer.
Which of the following is true?

- when new training data are provided to the neural network and training progresses, the units in the layer will continue to all have identical values for their weight parameters
- removing all units except one from the layer would result in an equivalent model
- at the start of training, initializing weight parameters randomly normally prevents the situation where all units in a layer have the same weight parameters at the end of training
- all units in the layer will have the same input and output as the other units in the layer
- all of the above

Which of the following is true about forward propagation?

- when evaluating a model, forward propagation is used to make predictions for data points, and these predictions are then compared to the points' true (known) labels
- as part of gradient descent, forward propagation is used to calculate the output of each layer, and these layer outputs are used in back propagation to determine gradients
- the results computed from forward propagation are used as part of the calculation of the cost of a model
- all of the above

Every layer of a neural network uses the same activation function.

TRUE FALSE

The *tanh* activation function outputs values between 0 and 1.

TRUE FALSE

Using no activation function is called using a “linear” activation function.
Using a linear activation function is common among hidden layers of a neural network.

TRUE FALSE

A neural network designed to solve a binary classificaiton problem should have two units in the output layer, with each unit predicting one of the two classes.

TRUE FALSE

A neural network designed to solve a multi-classificaiton problem should have *k* units in the output layer, with each unit predicting one of the *k* classes.

TRUE FALSE

All neural networks use the same loss function.

TRUE FALSE

Which of the following is likely to speed up execution of gradient descent?

- Start with a smaller value for the learning rate, α
- Use a larger network, e.g., more units per layer or more layers
- Use the regularization method of early stopping
- Increase the maximum number of iterations
- All of the above

When trying different architectures, each should be evaluated with the testing data before finalizing the choice of architecture.

TRUE FALSE

Overfitting occurs when a model learns properties specific to the training data that don't generalize well to new data.

TRUE FALSE

The main goal of regularization is to combat overfitting.

TRUE FALSE

Task 2: Coding with advanced neural networks

Download the Jupyter Notebook for Exercise 5 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

The above code should report that there are 4 units in the output layer. Why did **`sklearn`** use an output layer with 4 units?

Which executed more quickly, the model with a single hidden layer containing 50 units or the model with 5 hidden layers each containing 10 units?

In general, which yielded better performance on the validation data, the models with more units in a single hidden layer or the models with more hidden layers?

What accuracy do you observe on the test data?

How many diferent class labels are there for these data?

What activation function is used in the output layer of the model?

What activation function is used in the hidden layers of the model (consult the documentation here to find the answer)?

How many iterations of gradient descent did the best performing model require?

What accuracy do you observe on the test data?

How many units are in the output layer of the model?

What activation function is used in the output layer of the model?

For the model that performed best on the validation data, what R^2 value did it achieve on the test data?

For the model that performed best on the validation data, what R^2 value did it achieve on the test data?

What activation function is used in the output layer of the model?

For the model that performed best on the validation data, what score did it achieve on the test data?

CS344 Exercise 5 Final Page

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME
Exercise	