CS344 Exercise 6

Task 1: Deep learning frameworks

A ***dense*** layer of a NN is another way to say a ***fully connected*** layer.

TRUE FALSE

Automatic differentiation in deep learning frameworks obviates the need for users to explicitly program back propagation and gradient computation.

TRUE FALSE

A ***tensor*** refers to a unit in a layer of a deep neural network.

TRUE FALSE

In training NNs, the term ***adam*** refers to a commonly used cost function.

TRUE FALSE

In training NNs, an ***epoch*** refers to a pass through processing the training data.

TRUE FALSE

In a NN where the first hidden layer is fully connected, the number of parameters in the first hidden layer depends on the size (i.e., number of features) of the input.

TRUE FALSE

Regularization is used as a means for combating overfitting.

TRUE FALSE

Task 2: F1-score

Thus far, when evaluating the performance of a classifier, our measure of accuracy has been the percentage of classifications that the classifier makes correctly, i.e., if it classifies 35 out of 50 data points correctly then we say that the accuracy is 70%. This notion of accuracy may be reasonable when we have approximately the same number of examples in each class, but it can break down when we have imbalanced classes. For example, imagine we are classifying dollar bills as authentic (0) or counterfeit (1). Suppose we encounter many more authentic bills than counterfeit bills, e.g., 99% of bills are authentic and 1% are counterfeit. If we used a naïve algorithm for classification that *always* classified every bill as authentic, then it would have an accuracy of 99% because it would correctly classify all authentic bills (which account for 99% of all bills) even though it would incorrectly classify all counterfeit bills (which account for 1% of all bills). Such a naïve classifier achieves a high accuracy, 99%, even though it is not particularly useful. This is one motivation for using other measures (besides accuracy) of a classifier's performance, particularly when the classes are imbalanced.

A *confusion matrix* uses a table format to keep track of different aspects of a classifier's performance. For instance, in the binary classification problem of identifying authentic and counterfeit bills, the confusion matrix would keep track of:

- True Positives (TP) - the number of bills predicted to be counterfeit that are actually counterfeit
- False Positives (FP) - the number of bills predicted to be counterfeit that are actually not counterfeit
- True Negatives (TN) - the number of bills predicted not to be counterfeit that are actually not counterfeit
- False Negatives (FN) - the number of bills predicted not to be counterfeit that are actually counterfeit

Confusion matrix

	Actually Counterfeit	Actually Not Counterfeit
Predicted by classifier to be counterfeit	TP	FP
Predicted by classifier not to be counterfeit	FN	TN

Suppose for one thousand bills, a classifier predicts 30 to be counterfeit and 970 to be authentic. Of the 30 counterfeit predictions, 20 are actually counterfeit. Of the 970 authentic predictions, 965 are actually authentic. Fill in a confusion matrix below indicating the number of true positive, false positive, false negative, and true negative classifications.

	Actually Counterfeit	**Actually Not Counterfeit**
Predicted by classifier to be counterfeit		
Predicted by classifier not to be counterfeit		

Two common measures of a classifier's performance are *recall* (a.k.a., *sensitivity*) and *precision*. *Recall* is the percentage of actual positives (counterfeit bills in our example) that were correctly predicted to be positives (counterfeit). *Precision* is the percentage of predicted positives (predicted counterfeit bills) that are actual positives (counterfeit bills).

$$recall = \frac{\text{TP}}{\text{TP+FN}} \qquad precision = \frac{\text{TP}}{\text{TP+FP}}$$

In our example, high recall means that we are identifying a lot of the actual counterfeit bills. High precision means that when we predict a bill as counterfeit, our prediction is likely correct.

What is the recall and what is the precision associated with our counterfeit bill classifier?

It is desirable for a classifier to have high recall and high precision. However, there is often a trade-off between the two, where a classifier's recall can be improved at a cost to its precision, or the classifier's precision can be improved at a cost to its recall. For a given classifier, if you tweak its parameters so as to increase the number of positive predictions that the classifier makes (in our example, you cause the classifier to predict a larger number of bills as counterfeit), then which of the following are likely:

- Recall will increase and precision will decrease
- Precision will increase and recall will decrease

It is often helpful to have a single measure, rather than two measures (recall and precision), to evaluate a classifier. The F_1 score is a measure of a classifier's performance that combines recall and precision into one number. The F_1 score is defined as follows:

$$F_1 \text{ score} = 2 * \frac{recall * precision}{recall + precision}$$

What is the F_1 score for our counterfeit bill classifier above?

Instead, suppose we use a naïve classifier that predicts all one thousand bills are counterfeit even though only 30 are actually counterfeit. What is the F_1 score for this naïve classifier?

Task 3: Coding with TensorFlow

Download the Jupyter Notebook for Exercise 6 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.

Ionosphere

How many parameters does the model have?

On the validation data, what is the score (F1 or R^2) of the model after training for 100 epochs?

Emotion Turkish Music

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

What activation function is being used by the output layer of your model?

What loss function is being used for training?

How many parameters does the model have?

On the validation data, what is the score (F1 or R^2) of the model after training for 100 epochs?

Boston Housing

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

What activation function is being used by the output layer of your model?

What loss function is being used for training?

How many parameters does the model have?

On the validation data, what is the score (F1 or R^2) of the model after training for 100 epochs?

CS344 Exercise 6 Final Page

Name(s): ____

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	