Midterm

**Q1:**

The first chart is a classification model and it is a confusion matrix. It shows the result of the difference of the predicted label and the true label. There are four labels, the top-bottom axis shows the true labels and the left-right axis shows the predicted labels. We can learn from the chart that, for example, for label apple, the number of correct prediction is 4, which means that it is not misclassified, and it is the same as the label banana. While label grape has a high rate of misclassification. A perfect model has a dark blue diagonal that runs from top-left to bottom-right.

The second chart is a regression model and it is a lift curve. It shows two lines, one is expected and the other is predicted one. The x-axis is 0 to 100% of the dataset. The y-axis is ranged according to the values predicted. It is trying to tell us that how different is these two lines are. And the green line becomes above to the blue one and then falls down, the blue one becoming above the green line.

I will use accuracy to numerically measure the first chart because it is classification and use RMSE to numerically measure the second chart because it is regression.

**Q2:**

Age: encode\_numeric\_zscore

Favorite Color: encode\_text\_dummy

Length: encode\_numeric\_zscore

Gender: encode\_text\_dummy

It would be different if I were trying to use the value as input or as output.

When a value is used as input: we should fill all missing input with function missing\_median. Also use encode\_text\_dummy to encode textual/categorical values.

When a value is used as output: we will discard rows with missing output and use encode\_text\_index to encode textual/categorical values. Also output numeric values do not need to be encoded.

A particular zip code is related to a region so when encode a value like zip code it should be related to a range of latitude and longitude.

**Q3:**

In the real world, it is hard for us to get a number of test data by totally different training method and algorithm from the training one. So we have to use validation to get a better model and to generate better out of sample predictions. And k-fold validation will also be a better method to generate out of sample predictions.

Out of sample predictions are important because when it comes to a new data, and we have to predict this data with the model we choose. Moreover, we have to use out of sample predictions to compare with the true ones and evaluate the accuracy of the model we create and decide that to what extend we can trust the out of sample predictions when it comes a new data.

Overfitting is a phenomenon that we build a model perfectly fit the training data. Although we reduce the in sample error to zero, we cannot be sure it will also perfectly fit all new datasets. In many cases it may actually lead to the opposite effect. So we may not make our model too complicated, which may cause to overfitting. Validation and regularization are both used to cure the overfitting problem.

**Q4:**

Using too large of a **learning rate** may totally fail or may result a higher error than a lower learning rate.

Using too small of a learning rate will usually converge to a good solution, however, it may take a quite long time to process.

Common values for learning rate are: 0.1, 0.01, 0.001.

**Momentum** adds the scaled value of the previous weight change amount (νt-1) to the current weight change amount(νt).

A very common value for momentum is 0.9.

**Adam** estimates the first (mean) and second (variance) moments to determine the weight corrections. Adam begins with an exponentially decaying average of past gradients and this average accomplishes a similar goal as classic momentum update; however, its value is calculated automatically based on the current gradient. Adam is very tolerant to initial learning rate (η) and other training parameters.

Q5:

An example of a classification problem: according to patients’ symptoms to classify the disease and predict the new patient by his symptoms and decide what disease does the patient has.

An example of a regression problem: in a stock market, we want to predict the stock price, whether go up or down. So we can collect data of the price years ago and the feature data that relevant to the stock price such as the location of the company, income of the company, net margin of the company. With these dataset we can create a regression model that may fit.

For the example of a classification problem, I will calculate the classification accuracy and calculate classification log loss, and high accuracy and low log loss are wanted.

For the example of a regression problem, I will calculate the RMSE to evaluate the neural network, wanting to get a low RMSE.

Of course, I can also use cross validation to evaluate these two problems and get a out of sample error. Because we want this error to be small.