

Math 642, Introduction to Statistical Learning
Spring 2020
Final Exam

This exam is closed book and closed notes. The honor code is in effect.

Problem 1 (30 points): This problem is about Decision Trees

- a. When constructing a Decision Tree using recursive binary splitting, how do you decide which feature to start with?

We pick the feature that leads to the greatest possible to reduce RSS.

- b. Once you have a deep tree, how do you decide where to prune it?

Estimate the test error of each subtree using cross-validation methods to choose the cutoff point. One commonly used pruning method is cost complexity pruning. Apply cost complexity pruning to get a sequence of best subtrees and use k-folds validation on the subtrees to choose the tree size with the smallest cv error.

- c. The unpruned tree, Tree Model A, has 30 terminal nodes. The pruned tree, Tree Model B, has 20 terminal nodes. Which tree
- i. Is more flexible (A or B)?
A is more flexible
 - ii. Is more likely to exhibit significant bias (A or B)?
B is more likely to exhibit bias
 - iii. Is more likely to have the lowest training error (A or B)?
A is more likely has a lower training error
- d. Why do you use bagging when making a decision tree?

Bagging can reduce the high variance addressed by the tree model since the model could varies a lot because of data splitting.

- e. What is the primary drawback of bagging?

The primary drawback of bagging is that they start with the same feature and usually end up with the same first cutoff, which leads to the high correlation between the bagging models.

- f. How does random forest address the problem with bagging?

The random forest model creates each tree with randomly selected \sqrt{p} features out of p features, which solve the problem of high correlation addressed with bagging.

Continued Next Page

Problem 2 (35 points): This problem is about Classification using Logistic Regression and KNN

- a. The Sigmoid Function that we studied for Logistic Regression is $e^y / (1 + e^y)$.
What does the Sigmoid Function estimate in Logistic Regression?

The sigmoid function takes the log-odds and returns a probability between 0 and 1

- b. What is the value of the Sigmoid Function when

- i. $y = \text{Positive infinity?}$
1
ii. $y = \text{Negative infinity?}$
0
iii. $y = 0?$
0.5

- c. You are comparing two K Nearest Neighbor models, $K=1$ is Model A and $K=10$ is model B. Which of these two classifiers

- i. Is more likely to have a large bias (A or B)?

B is more likely to have a large bias

- ii. Is more likely to overfit the data (A or B)?

A is more likely to overfit the data

- iii. Is more likely to have the largest training error (A or B)?

B is more likely to have a larger training error

- iv. Is more likely to have the largest test error (A or B)?

Not sure whether A or B could have a larger test error. Since $K=1$ could be very flexible, I assume A would have a larger test error.

- d. Figure 1 shows a confusion matrix. From these numbers given in the confusion matrix, write the fraction representing the following (note: you don't have to calculate the number, just write the fraction)

- i. The Probability of Detection
 $81/(81+252)$
ii. The False Positive (or False Alarm) Rate
 $23/(9644+23)$
iii. The sensitivity
 $81/(81+252)$
iv. The specificity
 $9644/(9644+23)$

Continued Next Page

Problem 3 (35 points): In order to illustrate how neural networks function, you set up a small neural network that will function as an “OR” gate, shown in Figure 1.

- a. What would a possible set of weights be for the “OR” gate in Figure 1?

$$\beta_0 = -1$$

$$\beta_1 = 2$$

$$\beta_2 = 2$$

- b. BRIEFLY Define these terms relating to Neural Networks:

- i. Input Layer

The input features for the neural networks.

- ii. Hidden Layer

The layers of nodes between input layer and output layer, usually from 2 to L.

- iii. Activation Node

A neuron with an activation function. It receives inputs with associated weights and bias, and apply an activation function, such as sigmoid, relu and gives an output to the next layer.

- iv. Backpropagation

The backpropagation method calculates the gradient of the final layer of weights first and move backwards to update each of the weights in the neural network so that the cost function is minimized, and the actual output is closer to the target output.

- c. What is one advantage of a Neural Network model?

Neural Network model has the ability to handle a large number of features, and it's good for complicated classification problems that other classification methods perform poorly.

- d. What is one disadvantage of a Neural Network model?

The model is usually a black box and hard to interpret.

- e. What is a Convolutional Neural Network?

A convolution neural network is a neural network with one or more convolutional layer in the hidden layers. The convolution layer applies a filter to the input that detect certain features.

- f. What is one good application of a Convolutional Neural Network?

Image recognition and classification.

Have a Great Summer!

		Actual	
		Default	No Default
Predicted	Default	81	23
	No Default	252	9644

Figure 1: Credit Card Dataset for Problem 2 parts d

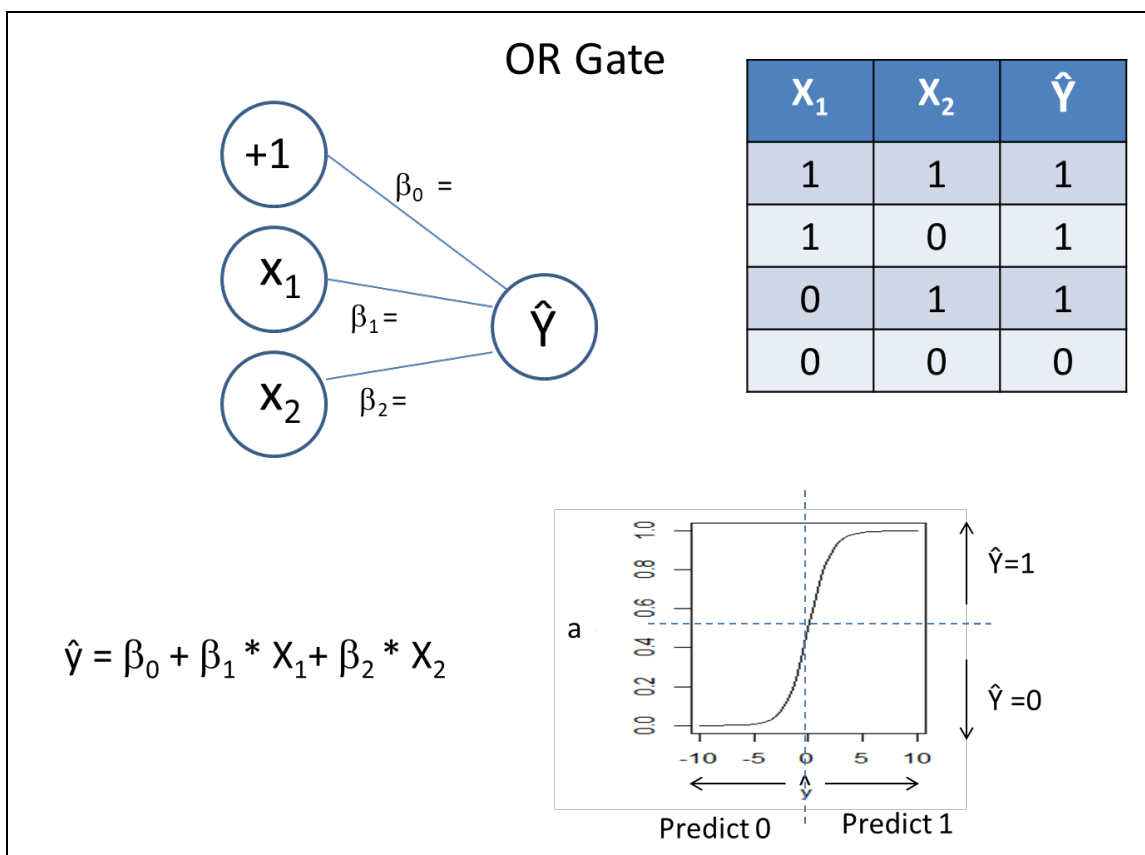


Figure 2: Neural Network “OR” Gate for Problem 3 part a

