Name: Ryan Mulligan

NETID: mullig47

Pledge: I have neither given nor received aid in this examination.

Signature:

CMSE381 QUIZ 1

Sep 24th, 2021

Instructions:
This is a closed book and closed notes examination. The best way to earn partial credits is to show all of your work. The instructor reserves the right to remove points if not all steps are shown. The total points are 50.You have 30 minus for this QUIZ. Good luck!

$f(x) = E(Y | X=x)$

1. [5 pts] Given two **continuous** random variable X and Y with joint probability density $f_{X,Y}(x, y)$. We seek a function $f(X)$ for predicting Y given values of the input X. If we use the squared error loss: $L(Y, f(X)) = (Y - f(X))^2$ to evaluate the model performance. Namely, we want to find a $f(X)$ to minimize the expectation of the squared error loss. What is the best function for predicting Y given $X = x$? (No proof is required).

5

The best function to do this would be to use the oracle function. The oracle function is $f(x) = E(Y | X=x)$. This function would help us minimize the expectation of the squared error loss because it shows us the best place to fit the data so that it isn't over/under fit.

2. [5 pts] Assume the same setting as previous question but now Y is a categorical random variable with three possible outcomes (tiger, elephant, and cat). We seek a classifier function $C(X)$ for predicting Y given values of the input X. If we use the misclassification error rate: $Error(Y, C(X)) = I_{(Y!=C(X))}$ to evaluate the model performance. What is the best function for predicting Y given $X = x$? (No proof is required).

2

The best way to be able to do this would be to create a model that takes in to accont all 3 animals, and then also takes into account differnt errors that may be in cluded. the model would look like

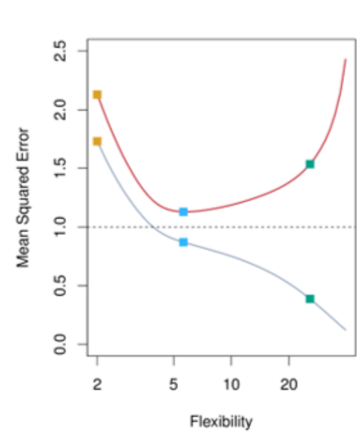
$$f(X) = \beta_0 + \beta_1 X + \beta_2 X + \beta_3 X + \epsilon$$

3. [5 pts] In a marketing setting, we have demographic information for a number of potential customers. We may wish to understand which types of customers are similar to each other by grouping individuals according to their observed characteristics. Is this a supervised or unsupervised learning? Explain your choice.

2

This is an example of Supervised learning because of the fact that We are getting Categorical data that we can Seperate groups into

4. The following figure displays the average training and testing MSEs as a function of model flexibility.



(a) [5 pts] Which curve is the testing MSE? Explain your choice.

5

The red line is testing MSE. We Can tell this because it dips down at the most ideal point, and then Continues back up after. MSE is Smallest when it has a good fit

(b) [5 pts] What is the meaning of the dashed line?

The dashed line is the area where error is unavoidable and the MSE can never go below because then it would not be taking error into account.

5

5. Assume the true model is

$$Y = f(X) + \epsilon.$$

We have a set of training data Tr, which is used to fit a model $\hat{f}(x)$, and a new testing data (x_0, y_0). Here, we assume x_0 are fixed. We have shown in the class that

$$E\left[\left(y_0 - \hat{f}(x_0)\right)^2\right] = Var(\hat{f}(x_0)) + [Bias(\hat{f}(x_0))]^2 + Var(\epsilon).$$

(a) [5 pts] Explain the meaning of $Var(\hat{f}(x_0))$ and $[Bias(\hat{f}(x_0))]^2$.

5

$Var(\hat{f}(x_0))$ - Variance of the function is how much it will differ from sample to sample randomly.

$[Bias(\hat{f}(x_0))]^2$ - How off the function is from what is to be expected, and how well the model is doing what it should be doing.

(b) [10 pts] A LASSO regression is to estimate $\beta = (\beta_1, \ldots, \beta_p)^T \in \mathbf{R}^p$ via

$$\text{argmin}_\beta \|Y - X\beta\|^2 \quad \text{subject to } |\beta_1| + \cdots + |\beta_p| \le \lambda,$$

3

where $Y \in \mathbf{R^n}$ and $X \in \mathbf{R}^{n \times p}$. Here, λ is the tuning parameter, which we need to specify. If we decrease the value of λ, how will the bias of the $\hat{\beta}$ change? Explain your answer.

If we Decrease λ, then the bias of the $\hat{\beta}$ will also go down. When we give it maximum constraints, there is no way for the bias to increase, so the only way it has options to go are either stay the same, or decrease

6. [10 pts] We are trying to predict the salary of workers from three states (Michigan, Ohio, and Indiana) using their age and residence using linear model. Write down the model and explain the meaning of the corresponding parameter (βs).

Same ?

4

$$f(X) + \mathcal{E} = \beta_0 + \beta_1 X + \beta_2 X + \beta_3 X + \mathcal{E}$$

This is the equation because we need to take into account each of the residents states. By using three different β values, we can assign each state a different β, and use that to predict their salary based on their location as well as their age.

7. (Extra 2 pts) For simple linear regression, we assume that $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$ and X is fixed (not random). We collect n i.i.d. training sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. Prove that the $(\hat{\beta}_0, \hat{\beta}_1)$ estimated through minimizing RSS equals to the one through maximizing likelihood.