Name:____________ PID:__________

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

Signature:____________

CMSE381 QUIZ 3 Maker-up

Nov 11th, 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!

---

1. We have a dataset $\{(x_1, y_1), \ldots, (x_{500}, y_{500})\}$ with sample size 500, where $x_i \in \mathbb{R}^{1000}$ and $y \in \mathbb{R}$.

a (5 pts) Can we directly fit a linear regression to the data? If so, what is the solution? If not, explain the reason.

b (5 pts) We want to performed a forward step-wise selection on this data set. Forward stepwise selection begins with a model containing no predictors, and then adds predictors to the model, one-at-a-time. In particular, at each step the variable that gives the greatest additional improvement to the fit is added to the model. Hence, we obtained K models $(\mathcal{M}_0, \ldots, \mathcal{M}_K)$ which contain $0, 1, 2, \ldots, K$ predictors respectively. Finally, we will select a single best model from among $\mathcal{M}_0, \ldots, \mathcal{M}_K$ using either AIC or BIC. If we want to have a sparse model (the number of predictors is small), which one should we use? Justify your answer.

2. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \quad \text{subject to } |\beta_1| + |\beta_2| \le s,$$

for a particular value of s.

a. (10 pts) Let β_1 be the $x-$axis and β_2 be the $y-$axis. Draw the area of (β_1, β_2) such that $|\beta_1| + |\beta_2| \le s$ and illustrate why this procedure will lead to a sparse solution when s is small.

b. (5 pts) As we increase s from 0, how will the training RSS change?

c. (5 pts) We collect a set of training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and let $\mathbf{Y} = (y_1, \ldots, y_n)^T$ and $\mathbf{X} = (x_1, \ldots, x_n)^T$, where $x_i = (1, x_{i1}, x_{i2})^T$. We then fit a model above with $s = 1$ to this training data. If we want to quantify the variance of the resulted estimate $(\hat{\beta}_1, \hat{\beta}_2)$, what method will you use (CV, LOOCV, Validation set, Boostrap, Hollycrap, MelowCP, or etc)? Briefly describe the procedure.

3. We want to build a MLP (multi-layer perceptron) neural network to predict the values of Y given $X \in \mathbb{R}^7$.

   a (5 pts) Before training the model, we normally need to preprocess the data. What are the major steps for the preprocessing?

   b (5 pts) Why do we normalize the inputs X?

4. Suppose your input is a 100 by 100 color (RGB) image.

   a. (5 pts) If you use a convolutional layer with 50 filters that are each 5×5 with padding size 2, stride of 2, and ReLu activation. How many parameters does this hidden layer have (including the bias parameters)?

   b. (5 pts) To avoid over-fitting, what strategy will you use? Briefly explain it.

5. (Extra 2 pts) Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\|Y - X\beta\|^2 + \lambda\|\beta\|^2,$$

where $Y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$. Derive the analytical solution of β.