ESE 5410 Midterm Practice Problems

Problem 1. Answer the following questions.

1. Explain whether the following scenario is a classification or regression problem. Finally, provide n and p: In order to predict housing prices in Philadelphia, we collect a set of data on 1000 houses. For each house we record number of bedrooms, number of bathrooms, and school scores. We are interested in understanding which factors affect housing prices.

2. Consider a binary classification problem for which you know that the conditional class probabilities are given by:

$$p_1(x) = \Pr(Y_i = 1 | X_i = x) = \frac{x^2 - \frac{1}{2}}{x^2 + \frac{1}{2}} \text{ and } p_0(x) = \Pr(Y_i = 0 | X_i = x) = 1 - p_1(x).$$

   Note that these probabilities are only defined in the range $x \leq -\frac{1}{\sqrt{2}}$ and $x \geq \frac{1}{\sqrt{2}}$. Find the optimal Bayesian classifier.

3. (True or False) In the medical case, the cost of False Negatives is much higher than the cost of False Positives.

Problem 2. Assume that 'nature' behaves according to the following linear model:

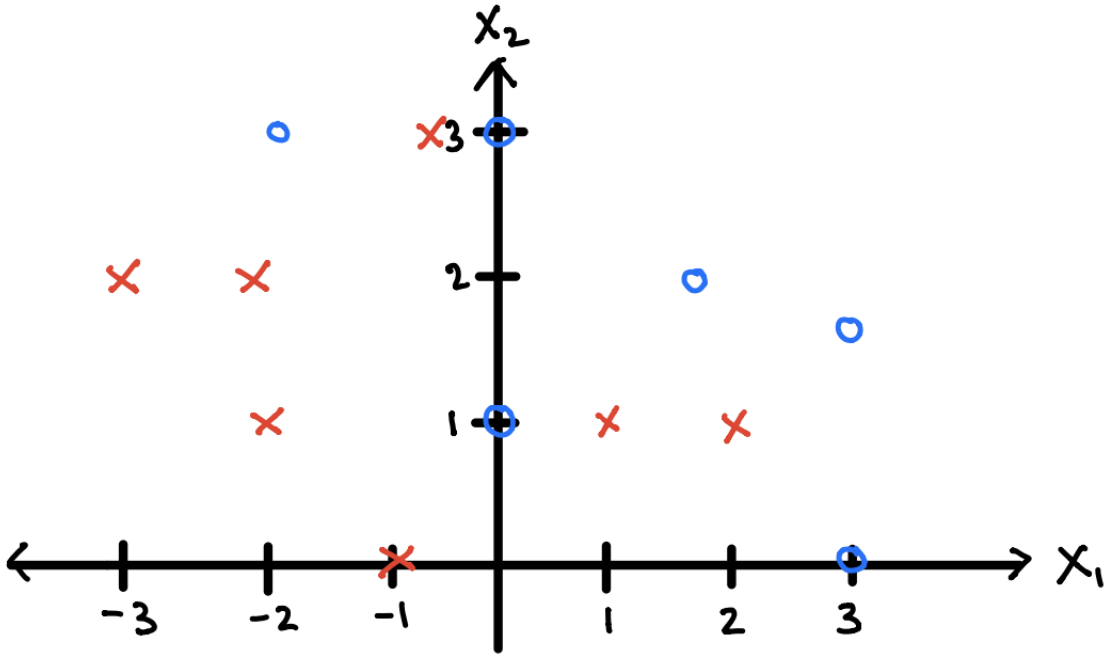
$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

where ε is a Gaussian random variable $\mathcal{N}\left(0, 8s^2\right)$.

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{5} = \{(-1, -3), (0, -2), (1, 0), (2, 2), (3, 3)\}.$$

1. Compute the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ for a linear estimator $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ using the data in $\mathcal{D}$.
2. Compute the RSS, RSE, and R^2 of the linear fit.
3. Compute the 95% confidence intervals for β_0 and β_1 as a function of s.
4. For what values of s would you reject the null hypothesis $H_0 : \beta_1 = 4$? Use a 95% confidence level in your answer.
5. For what values of s you *cannot* reject the null hypothesis $H_0 : \beta_0 = 2$? Use a 95% confidence level in your answer.

Problem 3. Consider the cloud of 13 points in the figure below, where the red points correspond to the label $Y = 1$ and the blue points to the label $Y = 0$. Assume we classify our points using a conditional class probability given by $g_1(x) = \Pr(Y = 0|X = x) = \sigma(2 + x_1 - x_2)$, where σ is the sigmoidal function ($\sigma(z) = e^z/(1 + e^z)$).



1. Draw in the x_1/x_2 plane the decision boundary associated with the Bayesian classifier, denoted by $C_B(\mathbf{x}) \in \{0, 1\}$.

2. Compute the number of the False Negatives, False Positives, and the confusion table of the sigmoidal classifier.

3. Consider another classifier $C_{NB}(\mathbf{x})$ satisfying $C_{NB}(\mathbf{x}) = 1 - C_B(\mathbf{x})$ (i.e., the new classifier flips the labels of the Bayesian classifier). Compute the number of the False Negatives, False Positives, and the confusion table.

4. Consider a classifier $C_2(\mathbf{x})$ where all outputs are 1, i.e., $C_2(\mathbf{x}) = 1$ for all x. Compute FNR and FPR of this classifier.

Problem 4: Answer the following questions (and make sure you justify your answers):

a. Consider a binary classification task for which you know that the posterior probability is given by

$$p_1(x) = \Pr(Y_i = 1 | X_i = x) = \frac{1}{1+x^2} \text{ and } p_0(x) = \Pr(Y_i = 0 | X_i = x) = 1 - p_1(x).$$

Find an explicit function for the Bayesian classifier $C(x)$, with the smallest error rate.

b. Explain, in simple words, what the FPR, FNR, and the TPR represent in the context of detecting spam email.

c. Describe, in simple words, a situation in which a variable X_2 is strongly correlated to another variable X_1, but X_2 is not directly caused by X_1. In general, can you infer causality from correlation?

d. Are independent variables always uncorrelated? Are dependent variables always correlated? Justify your answers.

e. Consider a variable $Y = X + X^2 + \varepsilon$, where $\varepsilon \sim \mathcal{N}(1, 2)$. Provide an expression for the regression function $f(x) = \mathbb{E}[Y|X = x]$. How much is the irreducible error?

f. Explain, in simple words, how you would use the MSE_{Te} to choose the optimal flexibility for a model.

g. *True* or *False*: A very flexible model can lead to overfitting when the Bias term in the Bias-Variance tradeoff is very high. Justify your answer.

h. Explain, in simple words, how to build the Bayes optimal classifier using the conditional class probabilities. Is it true that the Bayes optimal classifier always renders the smallest combined number of False Negative and False Positive errors? What about just the smallest number of False Negative errors?

Problem 5: Consider a *biased* coin with an *unknown* probability of heads, denoted by p.

a. Assume that we perform a coin flip experiment, tossing a coin $n = 3$ times, and obtaining the following sequence of Heads and Tails: H, H, T. What is the probability of observing that particular sequence of heads and tails as a function of the unknown variable p?

b. Assume now that we toss a coin 30 times. What is the probability of observing a sequence containing 20 heads and 10 tails (in any order) as a function of the unknown variable p?

Problem 6: Answer the following questions:

a. Suppose we have a data set with three predictors, X_1=GPA, X_2=Gender (1 for Female and 0 for Male), X_3=Interaction GPA×Gender. The response Y is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \times \text{GPA} + \hat{\beta}_2 \times \text{Gender} + \hat{\beta}_3 \left(\text{GPA} \times \text{Gender}\right),$$

and get $\hat{\beta}_0 = 35$, $\hat{\beta}_1 = 5$, $\hat{\beta}_2 = -15$, $\hat{\beta}_3 = 5$. Plot the two regression functions $\hat{Y}\left(X_1\right)$ corresponding to the cases X_2=Male and X_2=Female. Plot these two functions for $X_1 \in [0, 4]$.

b. For what range of values of GPA do Males earn more (on average) than Females?

c. Suppose that, using the same data as in Q2.a., we fit another predictor using the following nonlinear model:

$$\tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 \times \text{GPA} + \tilde{\beta}_2 \times \text{Gender} + \tilde{\beta}_3 \left(\text{GPA} \times \text{Gender}\right) + \tilde{\beta}_4 \left(\text{GPA}^2 \times \text{Gender}\right).$$

Is the training RSS for this model lower than the training RSS for the previous model.

d. If we compute the training RSS for $\hat{Y}$ and $\tilde{Y}$ using ONLY the data points corresponding to Males (discarding those of Females), which RSS is lower?

e. Would you answers change if we use the test RSS instead of the training RSS?

Problem 7: Assume that 'nature' behaves according to the following model:

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

where ε is a Gaussian random variable $\mathcal{N}\left(0, w^2\right)$. Using this model, nature generates a data set

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{5} = \{(-2, -3.5), (-1, -3), (0, 0), (1, 3), (2, 3.5)\}.$$

a. Compute the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ for a linear estimator $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ using the data in $\mathcal{D}$.

b. Compute the 95% confidence intervals for β_0 and β_1 as a function of w.

c. For what values of w can you reject the null hypothesis $H_0 : \beta_1 = 0$?

d. How would your answers change if the noise ε is a Gaussian $\mathcal{N}\left(1, w^2\right)$?

Problem 8: Consider a family of classifiers based on flipping a coin with Pr(Head) $= p$, such that we assign a label $\hat{Y} = 1$ if the coin comes out 'Head' and $\hat{Y} = 0$ if the coin comes out 'Tail'.

a. Prove that the ROC of this family of classifiers is a straight line.

b. What is the value of the AUC?

c. Without any knowledge about the input variable X, can you design a classifier achieving a higher AUC? How about a lower AUC? Justify your answers.