ECE-UY 4563: Introduction to Machine Learning
Midterm 1 Solutions, Fall 2021

Prof. Sundeep Rangan

- The exam is two parts:
  - Written part: This file, three questions
  - Python part: See the `Midterm1_2021_Python.ipynb` file. Also three questions.
- Submit answers to all six questions. You have at least 24 hours.
- You may use any resources on the web, class notes, and homework solutions. However, you may not ask friends or other classmates for help.
- Best of luck.

1. *Linear Regression.* Consider the model:

$$\widehat{y} = f(x) = \begin{cases} c_1 e^{-2x} + c_2 & \text{if } x < 1 \\ c_1 e^{-x} + c_3 & \text{if } x \geq 1. \end{cases} \tag{1}$$

for parameters $c = (c_1, c_2, c_3)$.

(a) Find **three** basis functions $\phi_1(x)$, $\phi_2(x)$, $\phi_3(x)$, such that every function in the model (1) can be written as:

$$\widehat{y} = \sum_{j=1}^{3} \beta_j \phi_j(x). \tag{2}$$

Write the β_j's in terms of the parameters c_j.

(b) Now find **two** basis functions, $\phi_1'(x)$ and $\phi_2'(x)$ for the model (1) with the additional constraint that $f(x)$ is continuous at $x = 1$.

Model	Training MSE		Test MSE	
	Mean	SE	Mean	SE
1	10	1	10.1	1.1
2	4	0.4	4.1	0.5
3	3	0.3	4.5	0.6

Table 1: Problem 2: Training and test MSE results for the three models

2. *Model Selection.* Consider three models for predicting a scalar y from $x = (x_1, x_2)$ where x_1 is a real valued variable and $x_2 \in \{1, 2, 3, 4\}$. Consider three models:

   - Model 1: $\widehat{y} = b + wx_1$ where w is constant and does not depend on x_2.
   - Model 2: $\widehat{y} = b + wx_1$ where w has one value when $x_2 = 1, 2$ and a different value when $x_2 = 3, 4$.
   - Model 3: $\widehat{y} = b + wx_1$ where w has one of four values depending on x_2.

   In all cases, the parameter b is constant and does not depend on x_1 or x_2.

   Answer the following with short explanations (e.g., one sentence) for each of the following:

   (a) Using a linear model for each case, what is the minimum number of parameters you need for each model?

   (b) Which model would generally give the lowest *training* error?

   (c) Which model would generally give the lowest *bias* error?

   (d) Which model would generally give the lowest *variance* error?

   (e) The results of K-fold validation are shown in Table 1. Which model would be selected based on the normal rule? Which model would be selected based on the one SE rule?

3. *Logistic Regression.* Consider a binary logistic classifier on a scalar x of the form

$$P(y = 1|x) = \frac{1}{1 + e^{-z}}, \quad z = b + wx.$$

(a) Find w and b such that

- $P(y = 1|x) = 0.8$ at $x = 3$ and
- $P(y = 0|x) = 0.8$ at $x = 1$.

(b) Using the values for w and b, what is $P(y = 1|x)$ at $x = 4$?