UIN: Name:

**CSCE-633 ML Midterm exam**

**Write your UIN and name at the top of each page. All 7 questions are equally weighted.**

Assume the following training set (3 training samples, each with 3 features) for a regression model:

$$x = \begin{bmatrix} \overrightarrow{x_1} \\ \overrightarrow{x_2} \\ \overrightarrow{x_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, y = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

1. What is the closed form solution for $w$ (provide a numerical answer) assuming Ordinary Least Square Linear Regression?

$$w = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

2. What is the closed form solution for $w$ (provide a numerical answer) assuming Ridge Regression with $\lambda = 1$?

$$w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

3. Assume the following training set (4 training samples, each with 2 features) for a binary classification model:

$$x = \begin{bmatrix} \overrightarrow{x_1} \\ \overrightarrow{x_2} \\ \overrightarrow{x_3} \\ \overrightarrow{x_4} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Denote the 2 features as $a$ and $b$. That is $\overrightarrow{x_i} = [a_i, b_i]$.

Suggest a new feature $c$ which is a function of $a$ and $b$ that, on its own, makes the training set linearly separable. Note: you may use only basic arithmetic (+,-,x,:,^,log). You are not allowed to suggest logic gates such as $c = \mathrm{XOR}(a, b)$.

$$(a - b)^2$$

4. After applying a kernelized SVM with a polynomial kernel (k=2) to a given dataset, you observe low training error yet high test error. What action is likely to reduce the test error? (select all that apply)

a. reduce the C hyperparameter towards zero

UIN: Name:

b. increase the C hyperparameter away from zero
c. use a polynomial kernel with K>2
d. add more IID training samples to the training set
e. remove all support vectors with alpha=0

5. When considering a decision tree model, how many distinct hypotheses exist for binary classification given a sample space of 2 features? Each feature has a discrete domain of 3 possible values.

$$2^{3^2} = 512$$

6. What is the information gain from splitting a data set of {6 true, 6 false} into three sets S_1={3 true, 3 false}, S_2={3 true, 0 false}, S_3={0 true, 3 false} ?

0.5

Alice trained a random forest model on three datasets $D_1, D_2, D_3$ drawn from a single training set $D$ of size $n > 3$. Select all statements that are necessarily true.

a) $D_1 = D_2$ implies that $\forall z, h_1(z) = h_2(z)$
b) $D_1 \neq D_2$ implies that $\forall z, h_1(z) \neq h_2(z)$
c) $D_1 \cap D_2 \cap D_3 \neq \emptyset$
d) $D \subseteq D_1 \cup D_2 \cup D_3$
e) Alice can use the "out of bag error" approach as long as $\exists i, D_i \neq D$
f) It is possible to have $D_1 = D_2 = D_3 \neq D$

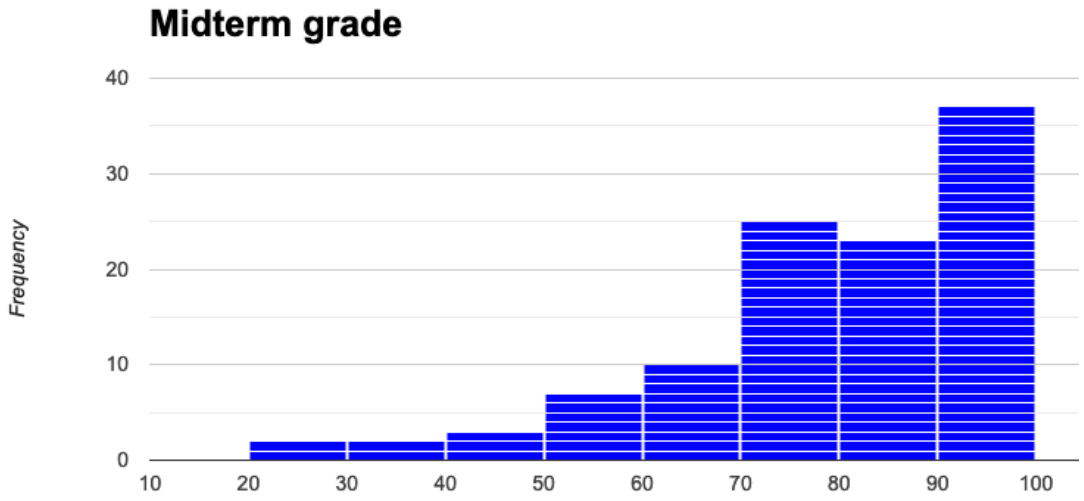
UIN: Name:

**Grading:**

Q (1)
14 points - Correct answer
7 points - Returning matrix (but w is diagonal(matrix))
2 points - If formula stated correctly
0 points - Otherwise

Q(2)
14 points - Correct answer
7 points - Returning matrix with w as the diagonal. scalar multiple of answer (i.e. not taking inverse, or dividing answer of Q1 by 4 (instead of 2)
2 points - If the formula stated correctly
0 points - otherwise

Q(3)
14 points - Correct answer
7 points - Abuse of notation in answer, but still understandable. for example (1 + a^T b) ^ 2. Note a*T b is scalar. but (1 + a.b) ^ 2 makes sense. Incorrectly using a, b as x1, x2 in answer.
4 points - Correct answer but using unpermitted operators like || or mod
2 points – Setting the new parameter not as a function of a and b
0 points - Otherwise




Average: 79.13