(1) In the following assignment, X denotes the data matrix of shape (d, n) where d and n are the number of features and samples, respectively.

(2) x_i denotes the i^{th} sample and y_i denotes the corresponding label.

(3) w denotes the weight vector (parameter) in the linear regression model.

1) An ML engineer comes up with two different models for the same dataset. The performances of these two models on the training dataset and test dataset are as follows:

Model 1: Training error = 0.9; Test error = 0.1
Model 2: Training error = 0.1; Test error = 10
Which model you would select?

- [x] Model 1
- [ ] Model 2

2) Consider a model h for a given d-dimensional training data points $x_1, x_2, ..., x_n$ and corresponding labels $y_1, y_2, ..., y_n$ as follows:

$h : \mathbb{R}^d \rightarrow \mathbb{R}$
$h(x_i) = \overline{y}$
where $\overline{y}$ is the average of all the labels. Which of the following error function will always give the zero training error for the above model?

- [ ] $\sum_{i=1}^{n}(h(x_i) - y_i)^2$
- [ ] $\sum_{i=1}^{n}|(h(x_i) - y_i)|$
- [x] $\sum_{i=1}^{n}(h(x_i) - y_i)$
- [ ] $\sum_{i=1}^{n}(h(x_i) - y_i)^3$

Common Data for Questions 3 & 4

Consider the following dataset with one feature

x_1	$label(y)$
-1	5
0	7
1	6

3) We want to fit a linear regression model of the form $y_i = w^T x_i$. Assume that the initial weight vector is $w = [2]$. What will be the weight after one iteration using the gradient descent algorithm assuming squared loss function? Assume the learning rate is $\eta = 1$.

-4

4) If we stop the algorithm at the weight calculated in question 1, what will be the prediction for the data point $x_1 = -2$?

8

5) Assume that w^t denotes the updated weight after the t^{th} iteration in the stochastic gradient descent. At each step, a random sample of the data points is considered for weight update. What will be the final weight w after T iterations?

- w^T
- $w^1 + w^2 + ... + w^T$
- $\frac{1}{T}\sum_{i=1}^{T} w^i$
- any of the w^t

Common data for questions 6 and 7

Kernel regression with a polynomial kernel of degree two is applied on a data set X, y. Let the weight vector be

$w = X[0.3, 1.6, 4.2, -0.5, 0.9]$

6) Which data point plays the most important role in predicting the outcome for an unseen data point? Write the data point index as per matrix X assuming indices start from 1.

3

7) What will be the prediction for the data point $[0, 0, 0, 0, 0]^T$?

6.5

8) If w^* be the solution to the optimization problem of the linear regression model, which of the following expression is always correct?

- $y^T X^T w^* = 0$
- $(y - X^T w^*)^T (X^T w^*) = 0$
- $(X^T w^*)(X^T w^*) = 0$
- $y - X^T w^* = 0$

9) The gradient descent with a constant learning rate of $\eta = 1$ for a convex function starts oscillating around the local minima. What should be the ideal response in this case?

- Increase the value of η
- Decrease the value of η

10) Is the following statement true or false?

Error in the linear regression model is assumed to have constant variance.

- True
- False