CS 550 Homework 1 (25 marks)

Instructions:

a. Due date is Aug 17. Hard copy has to submitted in class or as a document in a Google Form before 8:30 am.
b. Please type your solutions. We won't grade hand-written answers.
c. Only individual attempts and original answers will get you the credits. Copying will lead to 0 marks and penalties will be imposed.

1. (6 marks) Case Study of Smart Taxi Pricing:

Smart Taxi uses an online perceptron algorithm in their mobile app to price taxi fares in Rs. The major feature used for pricing is the length of the journey.

L**:** Length of the journey (kms)

The algorithm generates an estimated price and the customer may or may not purchase the ticket. Using the data below, write the equations of the proposed price generated by the algorithm.

Pricing Equation: P = C*max {L, 20} + 100

Starting values of C is 20. However, the value of coefficients C changes with time based on the demand/behavior of the users according to the following rule.

i. If user accepts the rides at the quoted price:

$$C = C + 0.01 * \nabla_c P$$

Where $\nabla_c P$ is the gradient of Price with respect to C

ii. If the user doesn't buy the ticket at the quoted price

$$C = C - 0.01 * \nabla_c P$$

Question a (1 marks): What is the minimum price of a ticket using the starting values of the coefficient?

Question b (2 marks): Explain what the above update rules for the coefficient C, is trying to do and comment if they are reasonable. Suggest a simple modification to the rule which stops this from getting very high values and avoid a situation like: https://www.nytimes.com/2022/04/12/nyregion/uber-lyft-surge-pricing-brooklyn-subway-shooting.html

Question c (3 marks): Please calculate the values of the coefficients after each transaction below and complete the table. Initial value of C =5.

L	Price Quoted	User Purchases?	C
10		Yes	
20		Yes	
10		Yes	
15		No	
35		Yes	
30		No	

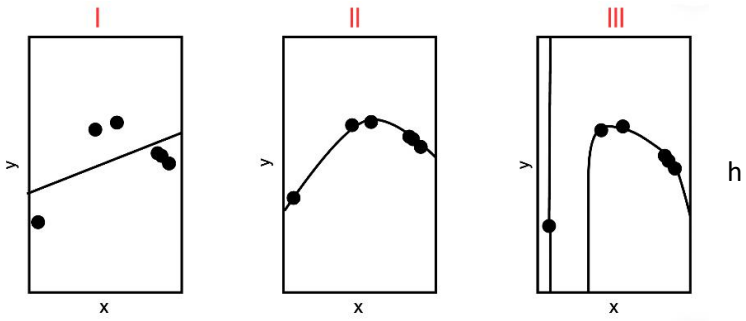
2. (10 marks) (1 mark each) **Objective type questions. Answer with a brief justification.**

i. **In machine learning, what is the purpose of splitting a dataset into training and test sets?**
   - **a.** The training set is for model training and the test set is for model validation.
   - **b.** The test set is for additional training after training with the training set.
   - **c.** The training and test sets are for training different models in parallel.
   - **d.** The test set is for pre-training the network before training with the training set.

ii. **Which of the following areas is not good for ML?**
   - a. Fluctuating environments, where the best model keeps changing, such as stock markets, pricing of gold etc.
   - b. Getting insights about complex problems with large amounts of data
   - c. Traditional computing problems such as inverting matrices, multiplying numbers etc.
   - d. Medical applications such as detecting tumors in brain scans

iii. **Given that N training samples are available to train a ML model. How many training samples are processed by stochastic gradient descent at each iteration of the optimization?**
   - a. one training sample
   - b. N training samples
   - c. k training samples, chosen at random, where 1<k<N/2
   - d. k training samples, selected in order, where 1<k<N/2

iv. **Which of the following depicts a model that underfits the data?**



v. Which Linear regression training algorithm can you use if you have a training set with millions of features?

vi. Suppose the features in your training set have outliers. Which algorithms might suffer from this, and how? What can you do about it?

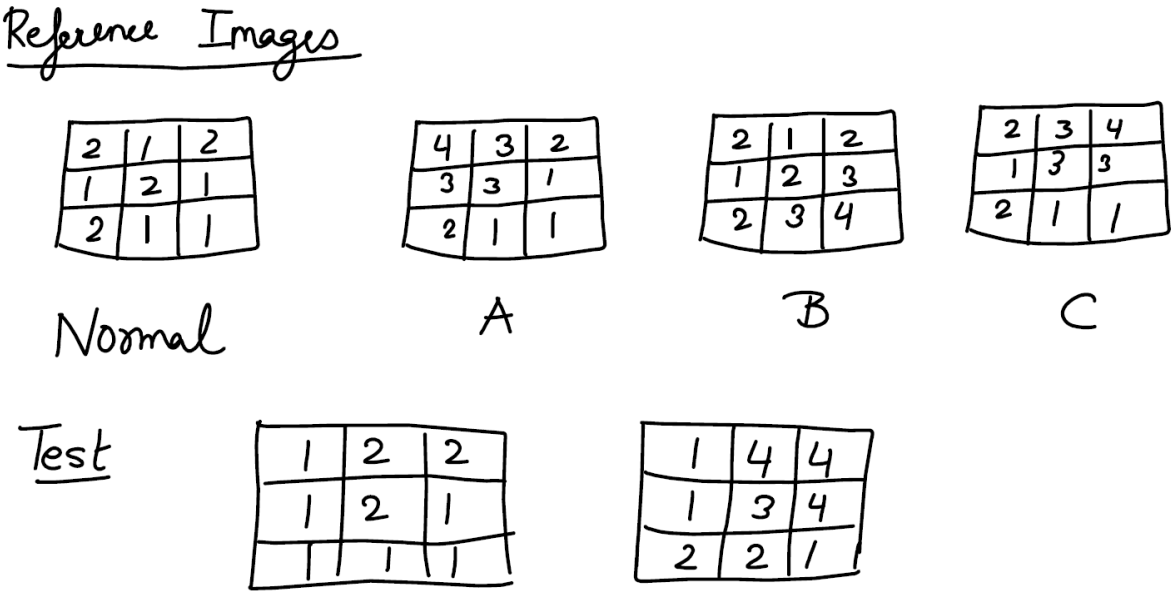
vii. Can gradient descent get stuck in a local minimum when training a Linear Regression model?

viii. Suppose you are using Ridge Regression and you notice that the training error and validation are almost equal and fairly high. Would you say that the model suffers from high bias or high variance? Should you increase the regularization hyperparameter λ or reduce it?

ix. Do all gradient descent algorithms lead to the same model, provided you let them run long enough?

x. When would you use Lasso regularizer?

Q 3. (4 marks) Images are commonly used in medical diagnosis. A doctor is using 4 reference images to classify a disease. Using a suitable distance function, classify the test images below.



Q4. (5 marks) In the lecture, the GD update equation was derived to be

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} + \eta_t \sum_{n=1}^{N} 2\left(y_n - \boldsymbol{w}^{(t)\top}\boldsymbol{x}_n\right)\boldsymbol{x}_n$$

Assume $N = 1$, and show that GD update improves prediction on the training input ($\boldsymbol{x}_n$, y_n), i.e, y_n is closer to $\boldsymbol{w}^{(t+1)\top}\boldsymbol{x}_n$ than to $\boldsymbol{w}^{(t)\top}\boldsymbol{x}_n$