Assignment 01

Instructions

1. Each assignment can contain both theoretical and practical questions.
2. Use LaTeX (preferred) or Word for theoretical question responses.
3. Practical questions are in the provided Jupyter notebook. Use Google Colab (Preferred) or Jupyter Notebook to complete questions directly in the Jupyter Notebook. Include code changes and reasoning in the Jupyter Notebook. Convert the Jupyter Notebook into an HTML page for submission.
4. Submit a PDF or Word file with reponses to theoretical questions, a Jupyter Notebook, and an HTML page (both files) with completed practical questions.
5. A 25% penalty applies to submissions on the first day after the due date, and a 50% penalty for submissions 24 to 48 hours late. No submissions will be accepted beyond 48 hours past the due date.

Theoretical Questions

Question 1

Considering the Vapnik-Chervonenkis (VC) Dimension, address the following three questions:

a. Explain the concept of VC Dimension with an illustrative example.

b. Imagine eight points positioned equidistantly on a circular rim. Can the VC Dimension for a triangle be 8 in 2D space? Justify your answer with an example. If the answer is No, specify the correct VC Dimension for a triangle in 2D space.

c. For any set of N points, a learning algorithm H can perfectly represent all possible ways of dividing these points into two classes for all values of N less than or equal to 2^d, where d is the VC dimension of algorithm H. Is this statement accurate? Substantiate your answer with clear reasoning.

Question 2

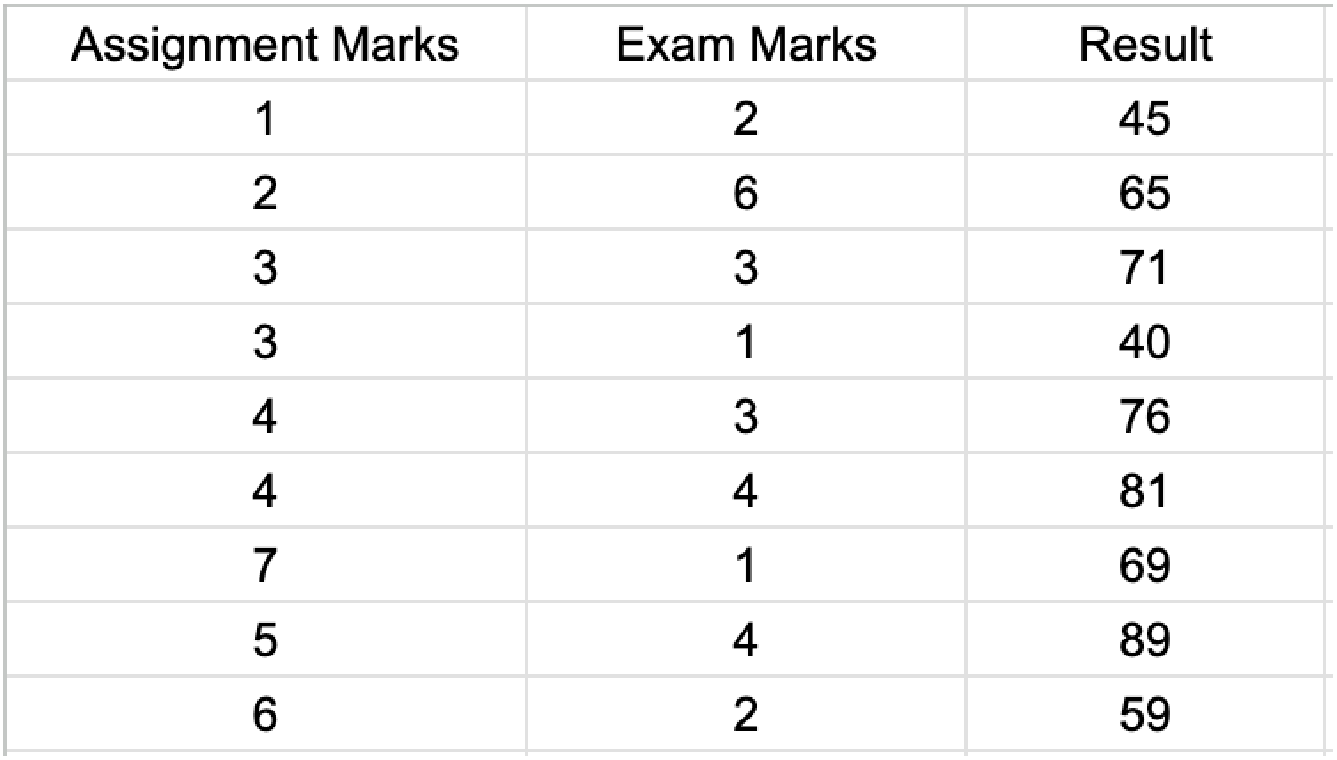
Consider the dataset below containing information about N students, including their Assignment and Exam marks along with corresponding results. The result is the target variable in the dataset. Assume you are running a regression model to predict the result, and the prediction formula is given as:

predicted_result = (0.83 * actual_result) + 15.

Calculate the predicted value for all the students using this formula and subsequently compute the error using the given formula:

$$Error = \frac{1}{N}\sum_{t=1}^{N}[predictedresult_t - actualresult_t]^2$$

Based on the calculated error, assess the performance of our regression model. Do you consider the model's performance to be good or bad? If you were to encounter the same level of error in general for other datasets, would it indicate the model's ability to predict data correctly across



Assignment Marks	Exam Marks	Result
1	2	45
2	6	65
3	3	71
3	1	40
4	3	76
4	4	81
7	1	69
5	4	89
6	2	59

Figure 1: Dataset Image

various cases? Justify your response. (Note: There is no absolute right or wrong answer; provide a reasoned explanation for your stance.)

Question 3

a. Explain the rationale behind the practice of dividing a dataset into Training, Validation, and Test sets. Specifically, elaborate on the advantages of incorporating a Validation set into our dataset for machine learning tasks.

b. Evaluate the decision to use the Training data as both the Validation and Test data. Provide a well-justified response, discussing the potential implications and drawbacks associated with such a choice.

Practical Questions

Please refer to and answer Question 4 and Question 5 in the provided Jupyter Notebook