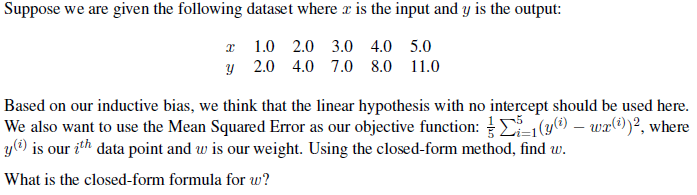
1.



We are trying to find the best fit line for a given dataset.

The objective is to find the weight ‘w’ that minimizes the Mean Squared Error (MSE) between the predicted and actual output.

The closed-form solution for a simple linear regression problem can be found using the formula:  
w = (X\*^T\*X)^{\*-1\*}X\*^T\*y, where:

(X) is the matrix of \*input\* data (y) is the vector of \*output\* data

Given this dataset, we can arrange your input ‘x’ and output ‘y’ in vectors and substitute them into this formula to find ‘w’. The weight that minimizes the MSE is ~ \*2.14\*

**Explanation :**

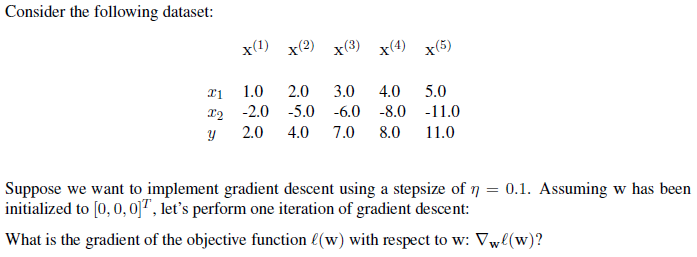
X^T = [

=> X^T.X = 1\*1 +2\*2+ 3\*3 + 4\*4 + 5\*5= 55

=> X^T.y = 1\*2 + 2\*4 + 3\*7 + 4\*8+ 11\*5 = 118

=> w = = = 2,14

2.



Given the dataset and assuming we are using a linear model with no bias term, the objective function ℓ(w) is the Mean Squared Error (MSE), defined as:

ℓ(w) = 1/N Σ(yi - w^T xi)^2

where:

N is the \*number of data points\*

yi is the \*actual\* output for the ith data point

xi is the \*input vector\* for the ith data point

w is the \*weight vector\*

The gradient of ℓ(w) with respect to w, ∇wℓ(w), can be calculated as:

∇wℓ(w) = \*-2\*/\*N\* Σ(\*yi\* - \*w^T\* .xi).\*xi\*

Given that w has been initialized to [0, 0, 0]T, we can calculate ∇wℓ(w) using the given dataset.

The update rule for gradient descent is:

w\_new = \*w\_old\* - η∇wℓ(\*w\_old\*)

where η is the step size. After calculating ∇wℓ(w), we can use this rule to update w.

If we weren’t given the step size, there are several strategies we could use to choose η. One common method is to try a range of values and choose the one that results in the \*lowest/smallest/minimal\* validation error. Another method is to use a technique like line search or learning rate decay, which adaptively adjust η based on how quickly the objective function is \*decreasing/decreased\*.

3.

Given a dataset with three input features (,,) and one output (y):

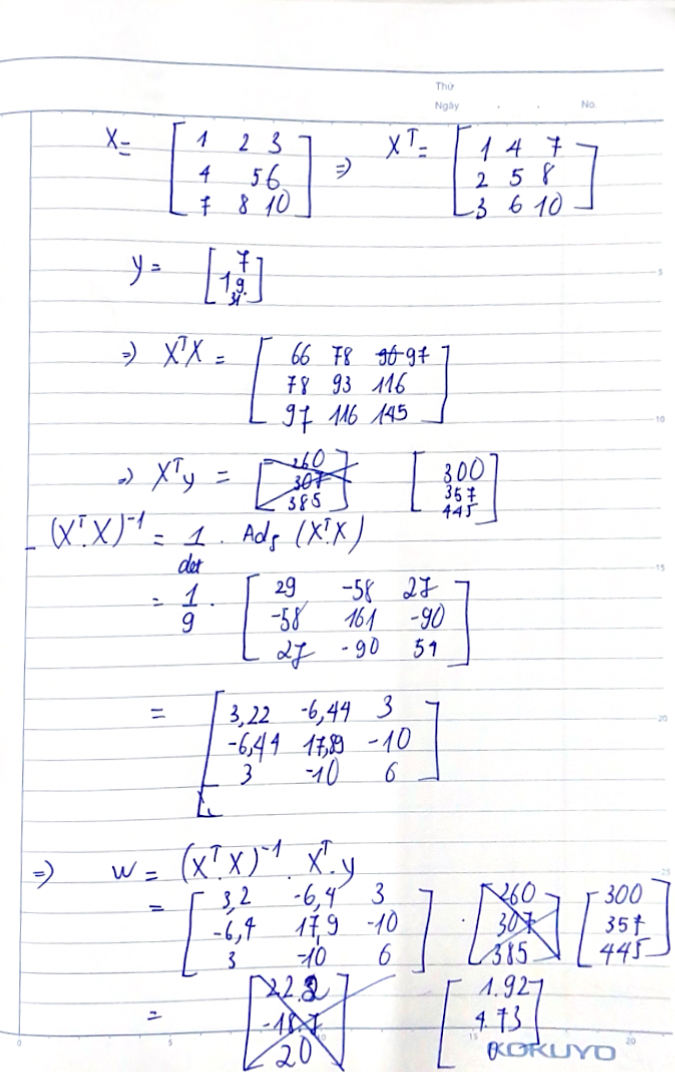
|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | y |
| 1 | 2 | 3 | 7 |
| 4 | 5 | 6 | 19 |
| 7 | 8 | 10 | 31 |

The weight vector w that minimizes the Mean Squared Error = **[1.92, 4.73, 0]**

Hint : The closed-form solution for the weight vector w in multivariable linear regression is given by:

w=(X^T.X)^(−1) \* X^T. y

**Explanation :**



Consider the following dataset with three input features and one output:

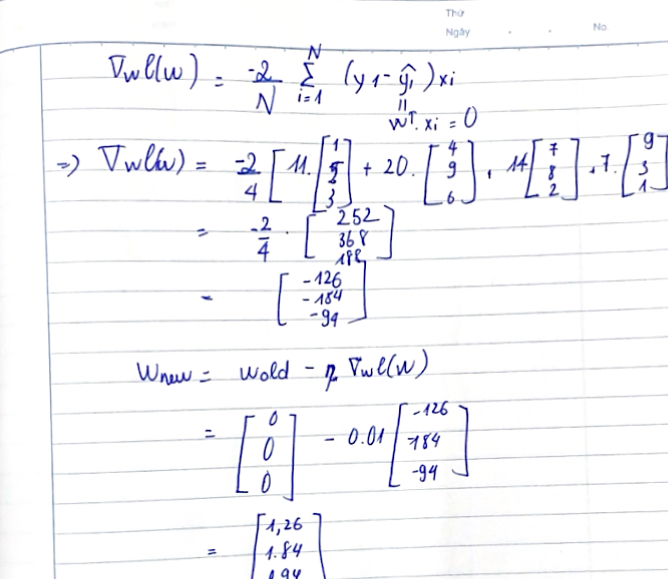
|  |  |  | y |
| --- | --- | --- | --- |
| 1 | 5 | 3 | 11 |
| 4 | 9 | 6 | 20 |
| 7 | 8 | 2 | 14 |
| 9 | 3 | 1 | 7 |

Initialize the weight vector w = [. Use a step size η=0.01.

Compute the gradient and update w to find .

**W\_new = [1.26, 1.84, 0.94]**

Explanation:



1. What is the main objective of the Bagging (Bootstrap Aggregating) technique in machine learning?

A) To increase the variance of the model

**B) To reduce the bias of the model**

C) To reduce the variance of the model

D) To reduce the data size

6. In Bagging, how are the individual models trained?

A) On the entire dataset

B) On a single sample of data

**C) On different subsets of the original dataset, created by sampling with replacement**

D) On different random subsets without replacement

7. Feature mappings in linear regression are used to:

A) Convert categorical data into numerical data.

**B) Increase the dimensionality of the input data to capture non-linear relationships.**

C) Reduce the number of features to prevent overfitting.

D) Randomly shuffle the data to avoid bias.

1. Which of the following is true about overfitting in a linear regression model?

**A) It occurs when the model fits the training data too well but performs poorly on new data.**

B) It occurs when the model has too few parameters and underfits the data.

C) It is always desirable as it minimizes the training error.

D) It can be avoided by increasing the complexity of the model.

1. Which of the following best describes vectorization in the context of linear regression?

A) Using for-loops to iterate through each data point individually.

**B) Representing computations in terms of matrices and vectors to improve computational efficiency.**

C) Adding more features to the model to increase its complexity.

D) Reducing the number of features to prevent overfitting.

1. In linear regression, the goal is to find the weight vector that minimizes the \_\_\_\_**MSE/mean square error**\_\_\_\_\_\_\_ between the predicted and actual outputs.
2. In gradient descent, the \_\_\_\_**learning rate**\_\_\_ controls the size of the steps taken towards the minimum of the loss function.
3. The \_\_\_**loss function** \_\_ is a function that quantifies the error between the predicted output and the actual output, guiding the optimization process.
4. In the context of linear regression, adding \_\_\_\_**a regularizer**\_\_\_ to the model helps prevent overfitting by constraining the size of the model parameters.
5. A **\_\_feature mapping\_\_** is a technique used to transform the original input features into a new space where a linear model might perform better.
6. When using gradient descent to minimize the loss function in linear regression, choosing a very small learning rate can sometimes result in oscillations around the minimum

**False.**

Explanation: A very small learning rate generally leads to very **slow convergence** rather **than oscillations**. Oscillations are more likely with a very large learning rate.

1. In linear regression with L2 regularization, increasing the regularization parameter λ will generally decrease the values of the model coefficients and make the model more biased but less variance-prone.

**True.**

Explanation: Increasing λ penalizes larger coefficients, which helps reduce variance but increases bias.

1. In the presence of collinearity among features, the closed-form solution for linear regression may become unstable and produce large coefficient estimates.

**True.**

Explanation: Collinearity can cause the matrix X^T.X to be nearly singular, leading to large variance in the coefficient estimates and making the closed-form solution unstable.

1. Regularization methods like L2 add constraints to the optimization problem but do not affect the structure of the model.

**False.**

Explantion: False. Regularization methods **change the structure of the model** by introducing penalties on the coefficients, **affecting their magnitude and selection**.