Assignment 3 (Linear Regression)

Problem Statement

You are given two separate train and test files. Compute the mean squared error of your result.

Coding Steps

1. Let your line be: $y_i = \beta_0 + \beta_1 x_i$

2. From the training file, compute (β_1 and β_0) using the following formula:

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

where SS_xy is the sum of cross-deviations of y and x:

$$SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}$$

and SS_xx is the sum of squared deviations of x:

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

where n = length of training dataset

3. For each input row in the test dataset, compute the difference between your output and given output (ε_i)

4. Compute mean squared error from the following formula and print your result.

$$\frac{1}{2n} \sum_{i=1}^{n} \varepsilon_i^2$$

where n = length of test dataset