TECHNOLOGICAL UNIVERSITY OF THE PHILIPPINES – MANILA
College of Engineering
Electronics Engineering Department
Ayala Blvd., Ermita, Manila, 1000, Philippines
Tel. No. +632-301-3001 | Fax No. +632-521-4063

Programming Assignment 2

LINEAR REGRESSION

In this assignment, you will train a model using Linear Regression. Go to https://archive.ics.uci.edu/ml/datasets/Energy+efficiency and download the Energy Efficiency dataset. The dataset contains 768 samples of simulated buildings with 8 attributes (X1 to X8) and two targets (Y1 and Y2). The following are their meanings:

X1 Relative Compactness
X2 Surface Area
X3 Wall Area
X4 Roof Area
X5 Overall Height
X6 Orientation
X7 Glazing Area
X8 Glazing Area Distribution

y1 Heating Load
y2 Cooling Load

The predictive model for heating and cooling loads of these buildings are useful for analyzing their energy consumption, in particular, in cold countries. In this programming assignment, the goal is to predict the heating load only

General Guidelines

1. Split the samples into *60% training, 20% validation, and 20% testing* data at random
2. Using Ridge regularization, set your own 10 different choices of regularization parameters, find the best choice that gives the highest accuracy on the validation data (based on R^2)
3. Make one final evaluation on the test data
4. What are the best model's coefficients, intercept, and its training, validation, and test accuracy?
5. What are the top 5 features among X1 to X8?
6. If you repeat the procedure above using only the 5 top features, what are the results?

Guide Questions

You are expected to answer the following questions using your analysis:

1. What is the impact of removing certain features on the model's performance?
2. Why is it necessary to split the data into training, validation, and testing sets? What could happen if this step is skipped?
3. How can the generalization ability of the model be assessed using the test data? What steps would you take to ensure the model generalizes well to unseen data?

Requirements

- Ensure that your code is clean, well-commented, and organized.
- Use Python libraries such as `numpy` and `pandas` for data manipulation and `matplotlib` or `seaborn` for visualization.

Submission

1. Submit your work as a Jupyter Notebook (.ipynb) file.
2. Upload your Jupyter Notebook to your GitHub repository. Ensure the notebook is well-documented with markdown cells explaining each step and the corresponding results.
3. Provide the link to your GitHub repository for grading.