DS413 - Introduction to Statistical Learning

Assignment-1
Assigned Date: 17/08/2024
11:59pm, Due Date:29/08/2024

Instructions

- Work on the assignments on your own. You are free to discuss among your selves,but don't copy. If we find the assignments of a group of (two or more) students very similar, the group will get zero points towards this assignment. Plagiarism will be cheked with tools. Please use Python for writing code. You can submit the code as a Jupyter notebook and For the theory questions, please submit your work to TAs. Use slack to discuss, if you have any doubts.

1 Programming Assignment

1.1 Regression

- **Mobile Phone Price Prediction** dataset. Implement the Linear Regression using the following methods. Use the matrices approach. Divide the data into train and test. Use training samples to construct data matrix and label matrix. Based on the dimension of data, assume the parameter matrix.

  1. Using closed form solution
  2. Using gradient descent
  3. Plot the predicted verses actual on test data for both the methods (1 and 2)
  4. Repeat the 1,2 and 3 with Regression with l_2 Regularization.
  5. Compare the results in 4 with and without data standardization.
  6. Plot the results from 4: predicted verses actual on test data by varying λ values
  7. After fitting the Regression with l_2 regularization, based on the weight values,comment on the importance of the features.

1.2 Linear Classification

- Fit a Linear Classification Model (2-class classification model) for the dataset given **Bank note authentication**: dataset

  1. Divide the dataset into training and testing
  2. Fit a classification model with and without l_2 regularization. Compare the classification accuracy on test data.
  3. Plot the test accuracy and training accuracy with various λ values. Write down your observations.
  4. Plot a 3D plot of your data by choosing important 3 features.
  5. Intentionally shift few data points and create outliers
  6. Fit Linear classifier on the dataset with outliers and what is impact of outliers on model's performance? comment?