Cong Liu A20343692

CS584 Assignment 1: Report

1. Problem statement
   Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. Linear regression is common term used in the field of statics. Linear regression is a foundational for machine learning area, and it also widely used in research area.
2. Proposed solution
   I try to increase factors of the features to see if the higher factors the better fittings. And it turns out that my assumption is wrong. Even though higher feature factors may fit the training data well, however, it can cause overfitting in training part. Sometimes higher factors would increase the variance. Our expected hypothesis error is the sum of variance error and bias error. Overfitting could increase the value of bias errors.
3. Implementation details
   One of the problems I met is trying to implement K-fold cross validation function. First, I should protect the raw data from being changed. Second, I should keep features and labels keep paired. Then, the more folds I implement, the more virtual memory is needed, and this would significantly increase the running time.
   To solve this problem, I use OO theory, encapsulate the functions and shuffle and cut data by index instead of do them directly on the data matrix. In this way, the security of data is protected and also increase the performance of my implementation.

4. Results and discussion
   Results and discuss are shown in the coding attachments.
   Demonstrations of correctness are also showed in attachments.

5. References
   [1]. Wikipedia: Linear Regression: https://en.wikipedia.org/wiki/Linear_regression
   [2]. Pattern Recognition and machine learning, Christopher M.Bishop,