Machine Learning - Assignment 01
Linear Regression

January 31, 2019

Problem

A Venture Capitalist Fund wants to see if there's any correlation between profit and the amounts that have been spent on different departments i.e Research and Development(R&D), Administration and Marketing, as well as in which state the company operates.

To test this idea, the VCF picks a sample of fifty startups randomly and records their Profit, R&D Spend, Admin Spend, Marketing Spend and State in which it operates.

It then runs a regression using:

- **Simple Linear Regression Model**
  - Cost Function : $\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2$
- **Ridge Regression Model**
  - Cost Function : $\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} w_j^2$
- **Lasso Regression Model**
  - Cost Function : $\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} \|w_j\|$

The optimization strategy used in all the above regression models is **Gradient Descent**.

Compare the accuracy of these models at learning rates(α) **0.005** and **0.01** and regularization parameter (λ) set as **0.05**.

Plot the graphs of all the regression models for the given set of learning rates and regularization parameter.