Introduction to Machine Learning
Lab Assignment 1: Regression

Due January 30, 2018 at 11:59pm

Summary: In class, we have discussed different regression models. For this assignment, you will demonstrate you understand how to apply, evaluate, and analyze these models.

1. **Comparative Analysis of Regression Models on Synthetic Data [40 points]**

   (a) Generate a dataset from a quadratic function with noise (similar to what we did in-class lab 2 with a linear function) that includes 1,000 samples. (Code)

   (b) Create a 80/20 train/test split of the dataset. (Code)

   (c) Train each of the following four regression models: linear, ridge, lasso, and polynomial (to the 2nd degree). (Code)

   (d) Report in a single table the predictive performance for each of the four regression models with respect to both of the following evaluation metrics: Correlation Coefficient and Mean Absolute Error. (Write-up)

   (e) Write a discussion analyzing and comparing the four models. For example, discuss which models you think are underfitting versus overfitting and explain why. Which method(s) perform the best/worst and why? Your discussion should consist of one to three paragraphs. (Write-up)

2. **Comparative Analysis of Regression Models on Real Data [40 points]**

   (a) Load any real dataset that is designed for the regression problem; e.g., Boston Housing data, Diabetes data. (Code)

   (b) Repeat steps 1b-1e for this real dataset.

3. **HyperParameters of Regularized Regression Models [20 points]**

   (a) **Ridge Regression**: describe the impact on the learned model when $\alpha = 0$.

   (b) **Ridge Regression**: describe the impact on the learned model when $\alpha = 1$.

   (c) **Lasso Regression**: describe the impact on the learned model when $\alpha = 0$.

   (d) **Lasso Regression**: describe the impact on the learned model when $\alpha = 1$.

How to Submit Lab Assignment 1: Please submit a pdf that provides hyperlinks to your code or answers to the questions, as deemed appropriate for the task. The pdf file should be named using your first and last name; i.e., firstname_lastname.pdf. The material you submit must be your own.