Machine Learning & Deep Learning
Homework 03

Fall 2019

1. What linear regression algorithm can we use if we have a training set with millions of features?

2. Suppose that the features in a training set have different scales. What algorithm might suffer from this characteristic and how? What can we do about it?

3. Why would we want to use:

(a) Ridge regression instead of linear regression?

(b) Lasso instead of ridge regression?

(c) Elastic Net instead of Lasso?

4. Is it a good idea to use mini-batch gradient descent immediately when the validation error goes up?

5. Do all gradient descent algorithms lead to the same model provided supposing we let them run long enough?

6. Using the Ames housing dataset (*www.amstat.org/publications/jse/v19n3/decock/AmesHousing.xls*), answer the following questions through a Jupyter Notebook. A description of the dataset is available at *jse.amstat.org/v19n3/decock/DataDocumentation.txt*

(a) Visualize the univariate distribution of each feature, and the distribution of the target. Do you notice anything? Is there something that might require special treatment?

(b) Visualize the dependency of the target on each continuous feature

(c) Split the data in training and test set. Then, evaluate the least mean square and visualize the relationship of the categorical variables that provide the best R^2 value with the target. Does scaling the data with `StandardScaler` help?

(d) Visualize the coefficients of the model. Do they agree with each feature's importance?