**CSCE 623 Spring 2020 - Machine Learning. In Class Work, Day 12**

From Chapter 6: Regularization - Ridge Regression & LASSO code practice:

Goal: To show the differences between ridge regression and lasso’s chosen coefficients – specifically how lasso can drive coefficients to zero while ridge regression cannot. You will be replicating many of the steps from lab 6.6 (pages 251-255), but you will augment the lab with additional steps. You will use synthetic data for this exercise.

You will need the sklearn, pandas, matplotlib, and numpy packages for this code

If you get stuck, visit <http://nbviewer.jupyter.org/github/JWarmenhoven/ISL-python/blob/master/Notebooks/Chapter%206.ipynb> for hints

1. You will generate datasets using the following function. Include this code early in your python code so you can call it later. Note that in this data, x is a predictor / feature and y is a response variable. Also note that this function is different from the one in your homework #3, so don’t reuse this code your homework code.

def makeDataQuadratic(myseed,quantity=100,myscale=0.0):

np.random.seed(myseed) #dont forget to import numpy as np

x = np.random.uniform(low=-1.,high=1.,size=quantity)

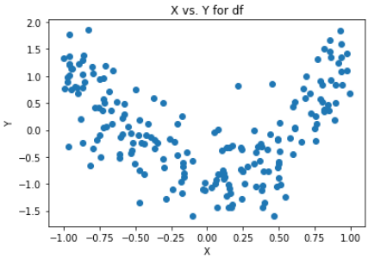
y = 2 \* (x \*\* 2)+np.random.normal(size = quantity,scale = myscale)

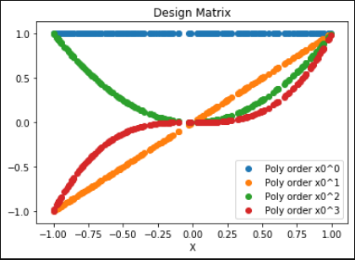
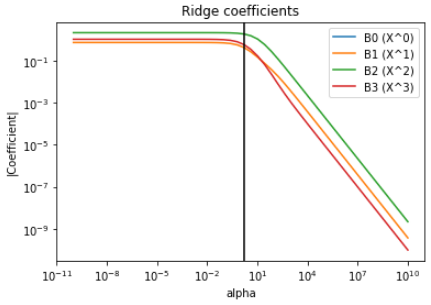
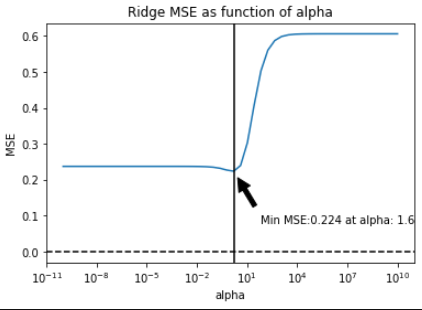
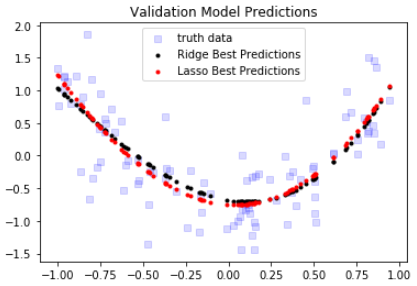
mean\_y = np.mean(y)

ydebiased = y-mean\_y #debias the data by subtracting the mean of y

df = pd.DataFrame({'x':x, 'y':ydebiased}) #dont forget to import pandas as pd

return(df)

Using this function, generate 200 datpoints with myscale=0.5. Scatterplot the data. Your data should look like this:

1. Using this dataset, build a set of predictors which includes x0 x1, x2 and x3. Use sklearn.PolynimialFeatures function – it generates a polynomial design matrix from a single feature vector x. The return matrix contains columns of x^0, x^1, … x^p where p is the desired highest order of the polynomial. Note that since it returns a design matrix, the columns correspond to *β*0 through *βp*  Using the dataset values for x, use the polyDesignMatrix function to generate a design matrix of order 3. Notice that the y is VERY dependent on x2. Thus in a well-fit model you would expect a large coefficient on a squared-x term and a smaller one on x.
2. Plot the values in the design matrix. You should get something like this:
3. Create a train/validation partition (using validation set method) as discussed in the lab 6.6 on page 253
4. Build a grid of possible alpha ridge regression penalty values (similar to lambda values discussed on page 251) in a logspace from -10 to 10
5. Use sklearn.linear\_model.Ridge to fit a ridge regression model and Collect the coefficients and the validation-set MSEs and predictions of yhat for each alpha. Display the coefficients and val-set MSE of the model at each value of alpha (known as lambda in the book on page 251 & 252). Do the coefficient values ever get to zero? (hint: they may be small… but test by comparing to zero). Describe how the MSE changes as a function of alpha. Why does MSE plateau at each extreme of alpha?
6. Make a loglog plot the absolute value of the coefficients of the ridge regression model as a function of the alpha value (page 251 & 252). Insert a horizontal black-dashed line at y=0 for reference (hint – use a loglog plot since your alphas are logarithmically spaced and the coefficents are probably covering multiple orders of magnitude). Insert a vertical black line at the alpha which minimizes the validation set MSE.
7. For ridge regression, make a semilogx plot of alpha value vs MSE on the predictions of the validation set. Find the index of the alpha value which minimizes MSE, report the alpha, report the MSE (optional – display these on the graph with a pointer and annotation)
8. Use sklearn.linear\_model.Lasso to train a lasso and collect the coefficients and val-set MSE and predictions of yhat of the model at each value of alpha, then display the coefficients and the MSEs of the model at the values of alpha (page 255)
9. Plot the coefficients of the lasso model as a function of the alpha value (page 251 & 252). What do you notice about the response of the lasso coefficients to the value of lambda?
10. For LASSO regression, make a semilogx plot of alpha value vs MSE on the predictions of the validation set. Insert a horizontal black-dashed line at y=0 for reference. Insert a vertical black line at the MSE-minimizing alpha.
11. Compare the minimum MSE models from the validation set performance on Ridge and LASSO, determine the best model and report it and its alpha value. Plot the truth data y values and the predictions from the best ridge and best LASSO models on a single graph