Alex Wolf
5/8/17

ML Regression Model Notes

**== to add still

Regression Modeling Steps/Notes

1. Things to check in Data
   a. if high dimensional
      -may want to use Lasso
   b. Check Initial feature correlation
      -**use scatterplot all feats function I came across earlier
      -Data should be correlated with Regression Label
      -Ridge Regression works well if highly correlated data
   c. ***get rid of any features with very low correlation with Label mainly, and other feats

2. Regression Model Testing
   a. Try RidgeReg as works well with good data
      c. Lasso first if very high dimensional Data or poor correlaton between all features
   b. Try Lasso Next
   c. Remove features
      - Plot Lasso Parameter coefficents and get rid of low features
   d. Try Ridge Regression again
   e. Try Elastic Net Regresion
   f. Test model fit versus number of parametes (AIC/BIC)
   g. Repeat if too many paramters from ideal AIC/BIC

Other Notes
   1. Alpha
      -high alpha= more regularization ==undefitting
      -Lower alpha= more ovefitting
      - normally from 0 to 1

   2. Lasso (L1 abs val_regualizaion)
      - good for achieving sparsity
      -difficult to avoid overfitting

-good for regression feature selection

3. Ridge (L1- square parameters)
   -Normally better bias/variance tradeoff
   -Good if Normal Prior distribution
   -Good for High dimensional Data
   -Much Better for Highly correlated features
   do correlation plot
   -Avoids overfitting more

4. Elastic Net
   - shrinkage and automatic variable reduction
   -find best combo of L1 and L2 regularization