Kaggle Notes:

* each x point has 16 data features
* Scoring Done with AUC, area under ROC curve
* Include your confidence level of that class value so that the ROC curve can be evaluated at different levels of specificity
* +1 = presence of rainfall

Plan:

* Which ones to do:
  + K-Nearest neighbor.
    - 5 models of K
      * get the AUC score
    - get the most efficient K with the highest score
    - Plot of the ROC curve
  + Linear models.
    - plot the ROC curve
  + Random forest.
    - plot the ROC curve

Adaboost:

* combining the output of weak algorithms into a weighted sum that represents the final output of the boosted classifier

<http://chrisalbon.com/machine-learning/random_forest_classifier_example_scikit.html>

Can use anything from sktools. But must explain the code and what you did. Must make report short.

HW 5 will be posted today

3/18/17

* Error rate from sklearn
* mean\_accuracy
* Kaggle performance score
* What to do:
  + Graph the ROC curve
    - use
    - graph either the error rate or mean accuracy or auc per data point
  + Validation vs Training curve
  + mean\_square\_error(y\_true, y\_hat)