**CS 386**

**Machine Learning for Watson**

**Project**

**Evaluation harness and final report**

**Due in stages**

**Final contest submission due on IBM site Friday, February 28, 10pm**

**Part Two due Sunday, March 2, midnight**

**Report due Saturday, March 8, midnight**

**Summary**

This is a continuation of the previous assignment. The goal of this assignment is to give you a better understanding of how to evaluate the classifiers that you are fitting to the IBM data, and perhaps climb the leaderboard as a result. We have created a skeleton for you to fill in for this project. Like the other homework assignments, this project can be done with your team. Note that if you run the code with the same options multiple times, you may get slightly different answers.

The bottom of this document tells you what to turn in – you will be answering some questions and writing some code.

**Instructions for the report will be available soon.**

**Instructions for Part Two:**

1. I have provided a new version of the code from part one, now called eval\_ibm3.py. It has similar functionality to the previous version, using methods from sklearn to create the cross-validation folds. This time, you will be filling in the compute\_pri method.
   1. You will fill in calculations for precision and recall. If you like, you can also add a calculation for the IBM score (optional!) (thus the name of the method: p=precision, r=recall, i=ibm score). See the class slides for these formulas.
   2. The inputs to the method are a list of the actual (correct) labels and a list of the labels returned by a classifier (predicted). The method returns three values for the results of the three calculations.
2. Run this to see what average P/R and IBM scores you get with a DecisionTreeClassifier, as for Part One (you can put this code in the same folder as for part one). Your scores may not exactly match those of part one, but should be reasonably close:

% python eval\_ibm3

* **What precision and recall did you get for cross-validation?**
* **What precision and recall did you get for stratified cross-validation?**
* **If implemented, what IBM score did you get for cross-validation and stratified cross-validation?**

You are of course invited to try this with other classifiers besides decision trees to see if you get a higher score!

**What To Turn In**

Turn in two files on Canvas (again, only one member of each team needs to turn in these files):

1. answers.txt, the answers to all questions asked above (in bold font). Provide the number of the question followed by the answer to the question
2. eval\_ibm3.py, your modified version of the code