Things to turn in:

1. A hard copy of write-up, submitted in class on Feb 2.
2. A zip file of the source code, mailed to kenghao@cs.berkeley.edu by Feb 2.

For this assignment, you should turn in a write-up of the work you've done, but not just the code (it is sometimes useful to mention code choices or even snippets in write-ups, and this is fine). We will grade your work mostly on your write-up, so you should focus on writing a report that is self-explanatory, and with sufficient details and decent comparison (remember writing a good paper is important for grad students). You should provide details in a way that audiences not taking this class could follow your write-up to reproduce results.

Important points to cover (but not limited to):

- The write-up should specify what models you implemented and what significant choices you made (e.g. Bernoulli or Multinomial models). You should explain how you chose a certain parameters (e.g. alpha for smoothing, systematically or by some random choice) and how that affected the performance of the system.
- It should include tables or graphs of the F1 scores, AUCs etc., of your systems.
- In particular, you should describe the term weights in your model, more specifically reporting the words with the top weights. You should include your reasoning of whether these words make sense and how you could possibly improve your result. You should describe what improvements you have done over the baseline naïve Bayes classifier.

We are more interested in knowing what observations you made about the models or data than having a reiteration of the formal definitions of the various models.

There is no set length for write-ups, but you can target about 3-4 pages, including your evaluation results, a graph or two, and some interesting examples.