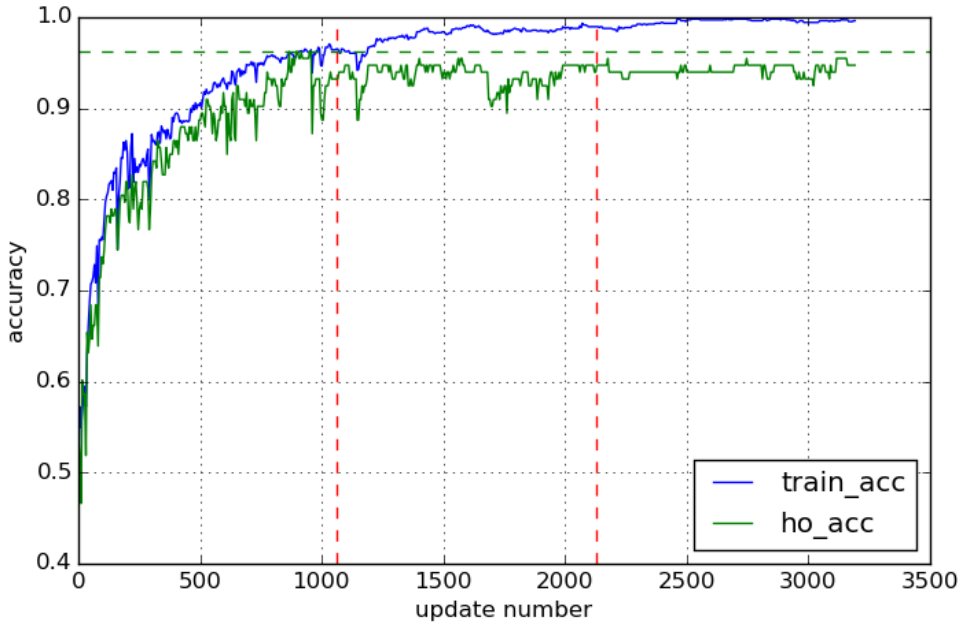
Name: Sanghee Kim
Subject: Logistic Regression
Submission date: Sat 7 Feb 2015 (use a free late day)

=== Analysis:

1. What is the role of the learning rate?
In stochastic gradient descent, another name of step size is learning rate.

2. How many passes over the data do you need to complete?



Just 1 pass is enough. Before the first pass(first red vertical line), test accuracy(ho_acc) reaches top accuracy. Also, after first pass(first red vertical line), I can see there has overfitting.

3. What words are the best predictors of each class? How (mathematically) did you find them?
If a word has the highest beta score, it should be the best predictor for positive class and vice versa. So, in this case, hit(1.20730829611) is the best predictor for baseball(positive), and hockey (-1.85194232277) is the best predictor for hockey (negative).

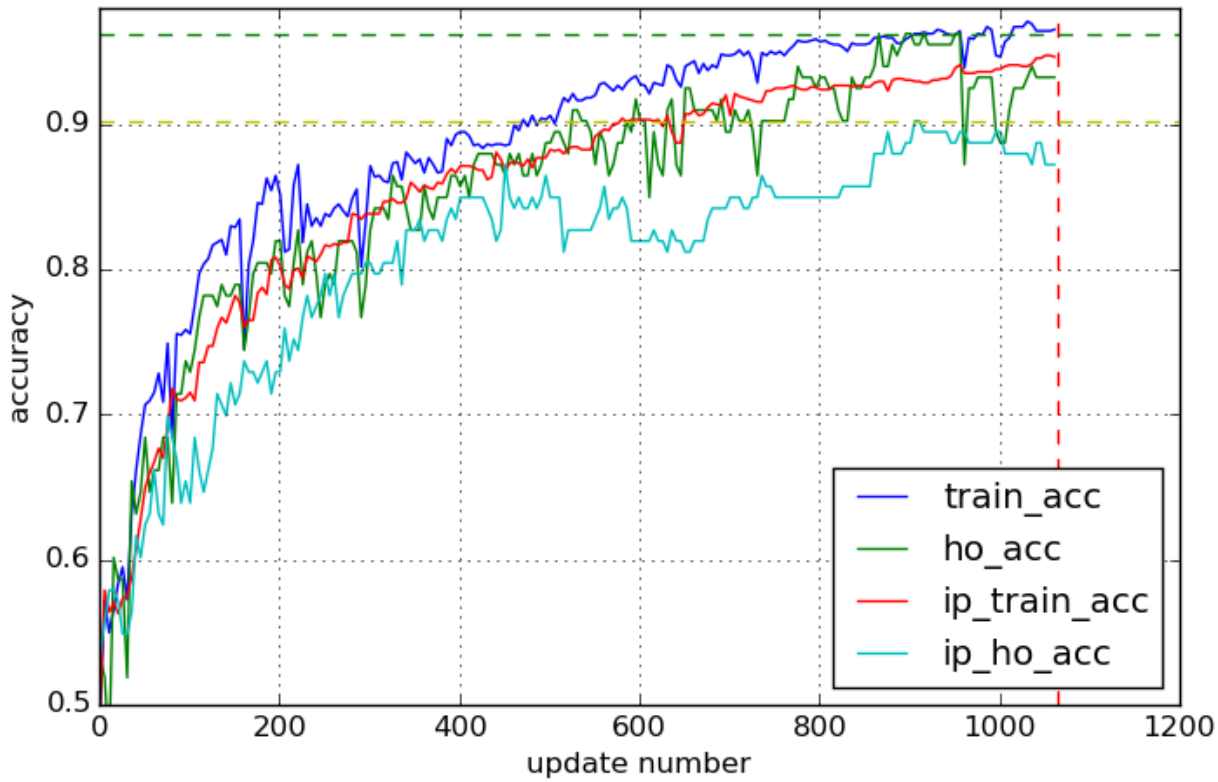
4. What words are the poorest predictors of each class? How (mathematically) did you find them?
If a word's beta score is 0, it mean that it doesn't help to predict class. I found 17 words which have 0 beta score. For example,

everywhere 0.0 blasted 0.0 intermissions 0.0 bloody 0.0 broad 0.0	deceased 0.0 hesitate 0.0 hooked 0.0 memoriam 0.0 pitiful 0.0	racist 0.0 riel 0.0 rode 0.0 silence 0.0 tone 0.0	vintage 0.0 wrestling 0.0

(Next page for extra credit)

== Extra credit:

1. Use a schedule to update the learning rate and show the effect.



For making learning rate which is a decreasing function of time, I made a function that is inversely proportional to the iteration(update number). *2.0 / iteration^0.5* is the function. In this case, ip_train_acc and ip_ho_acc were applied by this learning rate.

However, it negatively affected train and test accuracy, even it seemed that the accuracy of ip_train_acc and ip_ho_acc is higher than the others at the very beginning. (1st to 25th iteration)

Reference: http://users.ics.aalto.fi/jhollmen/dippa/node22.html

2. Use document frequency (provided in the vocabulary file) to modify the feature values to tf-idf and show the effect.

-