Machine Learning - CSCI 5622

HW 2 - LogReg

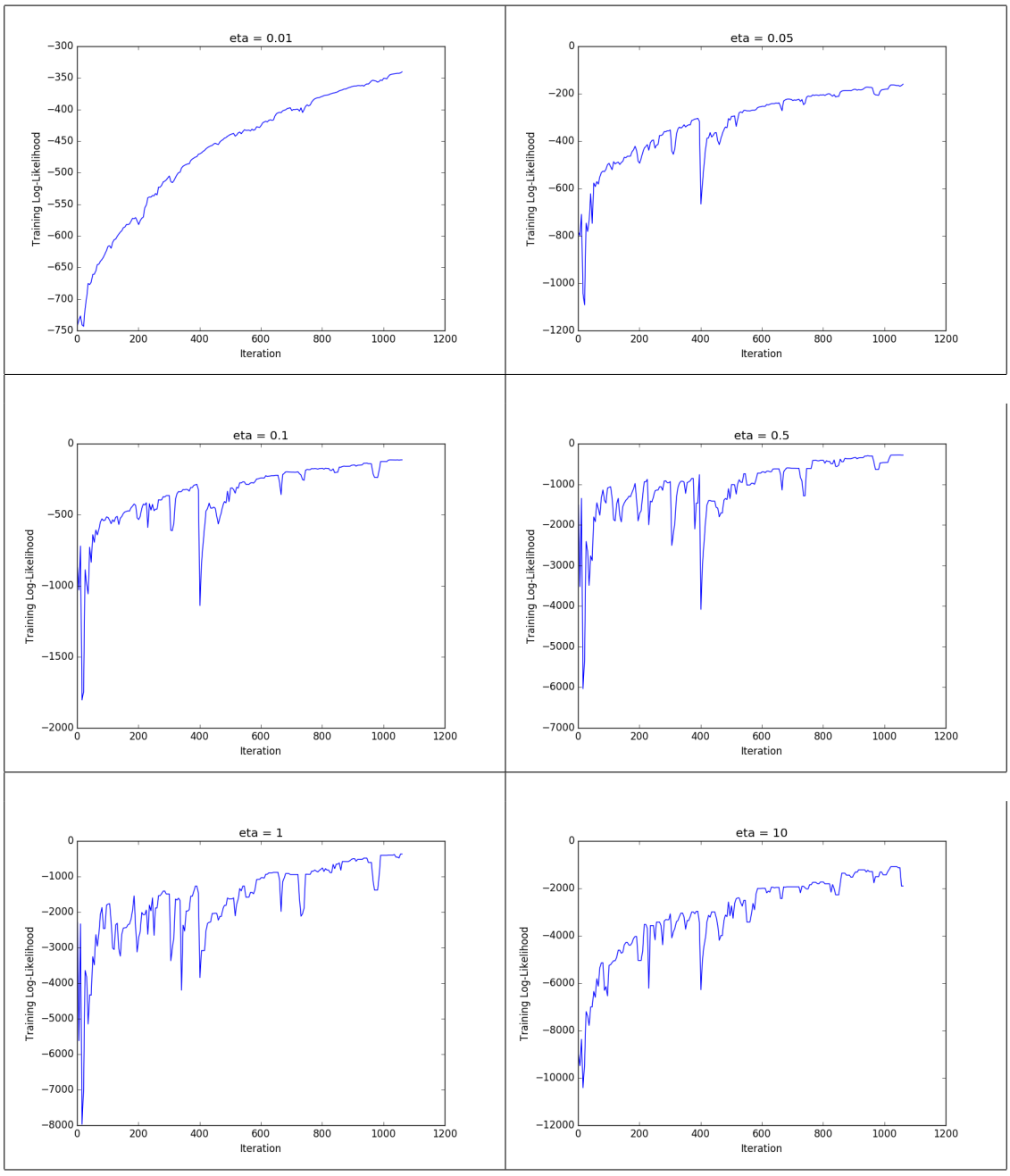
Santhanakrishnan Ramani

Friday 9th September, 2016

Analysis

1. How did the learning rate affect the convergence of your SGA implementation?

   The figure below shows the graphs of iteration vs training set log probability for different values of eta. We can see clearly that when eta value is very low the rate of convergence is slow, and when eta value is large it diverges and has a very low log probability. So the optimum value for eta will be somewhere near 0.1.



2. What was your stopping criterion and how many passes over the data did you need to complete before stopping?

   From the below table we can see the Holdout Accuracy increases until pass 8, then decreases. The stopping criteria is when the Holdout Accuracy stops increasing. The number of passes completed before stopping is 8.

Update 1064	TP -114.114816	HP -30.655089	TA 0.969925	HA 0.924812
Update 2128	TP -76.817798	HP -25.634843	TA 0.986842	HA 0.924812
Update 3192	TP -50.420317	HP -25.067925	TA 0.996241	HA 0.932331
Update 4256	TP -41.379968	HP -23.919768	TA 0.996241	HA 0.932331
Update 5320	TP -35.689365	HP -24.063290	TA 0.999060	HA 0.932331
Update 6384	TP -31.632305	HP -23.576720	TA 0.999060	HA 0.939850
Update 7448	TP -28.428944	HP -23.526622	TA 0.999060	HA 0.939850
Update 8512	TP -25.960753	HP -23.437327	TA 0.999060	HA 0.939850
Update 9576	TP -23.818219	HP -23.509580	TA 0.999060	HA 0.932331
Update 10640	TP -22.065836	HP -23.211715	TA 0.999060	HA 0.932331

3. What words are the best predictors of each class? How (mathematically) did you find them?

   - Best Predictors for Baseball: pitch, rotation, day, saves, run, baseball,hit
   - Best Predictors for Hockey: pts, hockey, period, vs, shots, playoff, ice, broadcast

   I choose words with high positive weights for Baseball, and high negative weights for Hockey.

4. What words are the poorest predictors of classes? How (mathematically) did you find them?

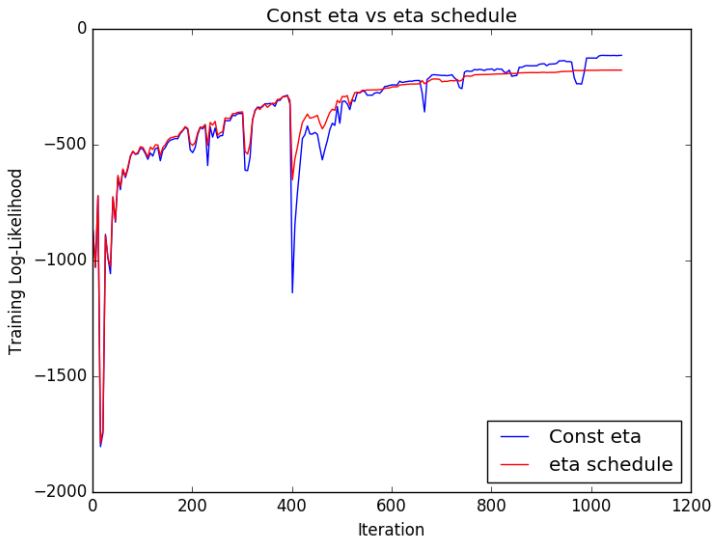
   - Worst predictors for both classes:
     contained, sportscasters, favour, prototypical, aggravatingly, pronunciation, crumbled, ratio, baby

   I chose the words that have either 0 or near 0 weights.

Extra Credit

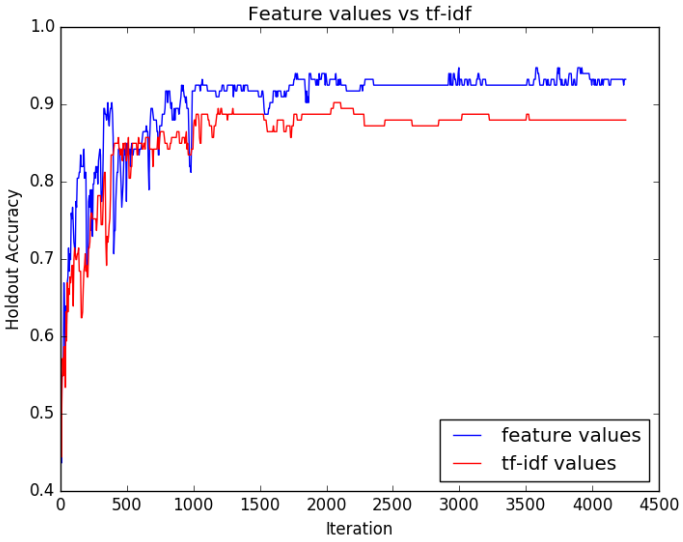
1. Using a schedule to update the learning rate and its effect.
   (use the argument –schedule 1 to use eta_schedule function)

   The figure below shows the plot of Iteration vs Training Log Likelihood while using constant eta and a eta schedule. It's clear from the graph that while using a eta schedule the convergence is faster and smoother.



2. Using document frequency to modify the feature values to tf-idf and its effect.
(use the argument –tfidf 1 to use ti-idf values)

From the figure below we can see that while using tf-idf the Holdout Accuracy kind of remains constant.



Accuracy of Holdout data when using feature values - 0.924812
Accuracy of Holdout data when using tf-idf values - 0.887218

So in this problem using tf-idf instead of feature values doesn't improve the accuracy of classification.