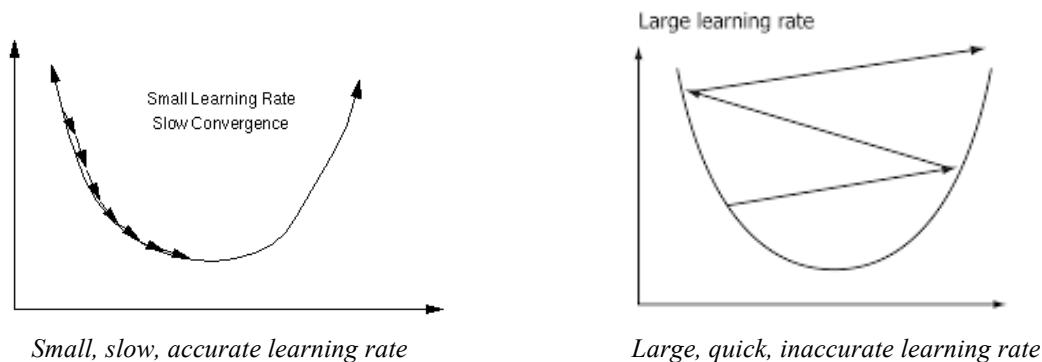


Jessica Petty
CSCI 3022
Homework 5
December 1, 2016

1. With each set of variables or features, there is some projection of how they will affect the classification. The learning rate determines how big of an affect we give a current projection. The concept of learning rate is very similar to the concept of step size in Euler's method in which a step is taken in the direction of the derivative and then another derivative is calculated. The larger the learning rate, the faster (and more inaccurate) the conversion. Similarly, the smaller the learning rate, the more slowly the algorithm converges, but the conversion is more accurate the classification.



2. I needed to complete 1061 passes over the data before the algorithm converged.
3. The best predictors of the classification were (1) hockey, (2) runs, and (3) catcher. These words are the words that had the greatest affect on the classification of the data. To find this, I simply found the largest magnitudes of values found in the beta vector that gives the weights (or importance) of each feature in determining a classification.
4. There were many features of this feature set that had no affect on classification. We see this when the value associated with this feature in the beta vector was zero. These features that had no affect are:
everywhere, blasted, hurled, intermissions, bloody, broad, chop, deceased, hesitate, hooked, memoriam, pitiful, racist, riel, rode, silence, tone, vintage, wrestling

Resources used:

<http://ufldl.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/>
<http://cseweb.ucsd.edu/~elkan/250B/logreg.pdf>

Additionally, Justin Schiller and I worked on developing the algorithm together for `sg_update`, though both our implementations came separately.