

Analysis

1. The learning rate is the increment distance of each step size. It helps us determine the rate when we modify the beta values of our function. While doing stochastic gradient descent our poll optimizes the objective function for our logistic regression model to find the maximum. This is done by computing the gradient of our function and then using it to find the slope. With the slope we know which direction to move on the objective function.

The learning rate helps to determine the step size when incrementing our beta values while trying to find the maximum. A small learning rate is more accurate and will produce a closer estimate to the actual maximum. This is also true when we see that a step size that is too large will possibly step over the actual maximum.

2. The number of passes depends on the value that beta updates to after the `sg_update` functions runs. This depends on the data set that we are working with. Generally we know to stop running and re updating the values once the beta values stop changing. We know we are at the maximum when the beta value converges and stops changing. At that point we can stop running further passes of `sg_update`.

3. The best predictors are the max and min beta values.

The best predictor for baseball is the word “run” with index 77 and beta value 1.2734

The best predictor for hockey is the word “hockey” with index 33 beta value -2.1096

4. The worst predictor for baseball is the word “everywhere” with index 2539 and beta value of about 0.