1. The learning rate helps determine the speed at which we approach the optimal weights. If it is too large then we will skip the optimal solution, as shown by the green arrows on the right. If it is too small, it will need too many iterations to converge on the best values, as shown by the blue arrows on the right. So a good learning rate finds a balance between being large enough to find the optimal solution in a manageable number of iterations, and not being so large that it keeps "overshooting" it.
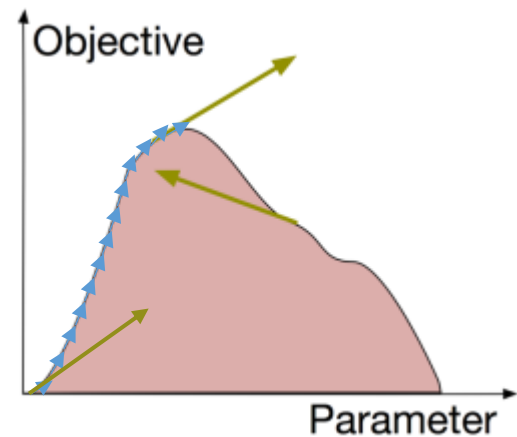
2. I will need to complete one pass over the data with 1064 updates. This is because after the first pass the beta values converge, therefore I can stop running more passes since the values don't update anymore.

3. Best Baseball Predictors:
    1. Runs
    2. Baseball
    3. Hit
    4. Pitching
    5. Catcher

   Best Hockey Predictors:
    1. Hockey
    2. Playoffs
    3. Golchowy
    4. Goals
    5. Pick

   I found these words based on their beta values. The largest positive values in my feature vector are strong predictors for baseball, while the largest negative values in my feature vector are strong predictors for hockey.

4. Poorest Predictors:
    1. Everywhere
    2. Bloody
    3. Blasted
    4. Vintage
    5. Hesitate

   These words are poor predictors of these two classes because they have beta values at or close to zero.