Analysis

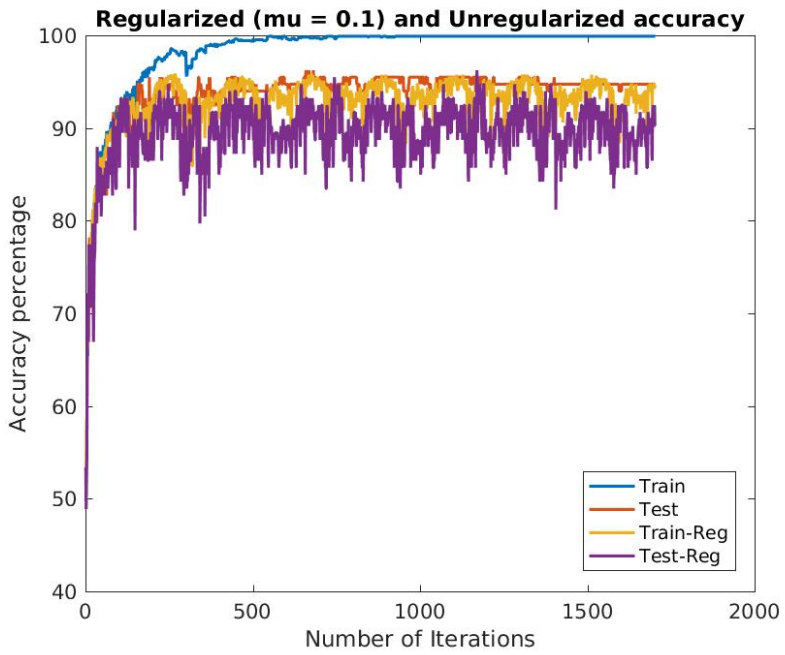


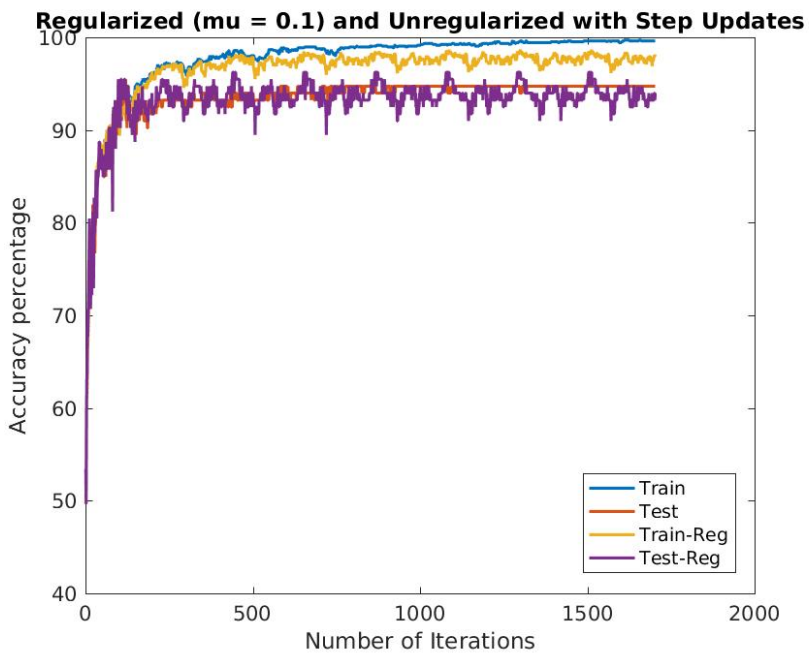
Figure 1: Normal Regularized and Unreglarized

Answer 1: The learning rate determines **how big of a step to take in the direction of the gradient**. A smaller than optimum rate can converge slowly and a larger than optimum rate can overshoot and oscillate around the minima/maxima.

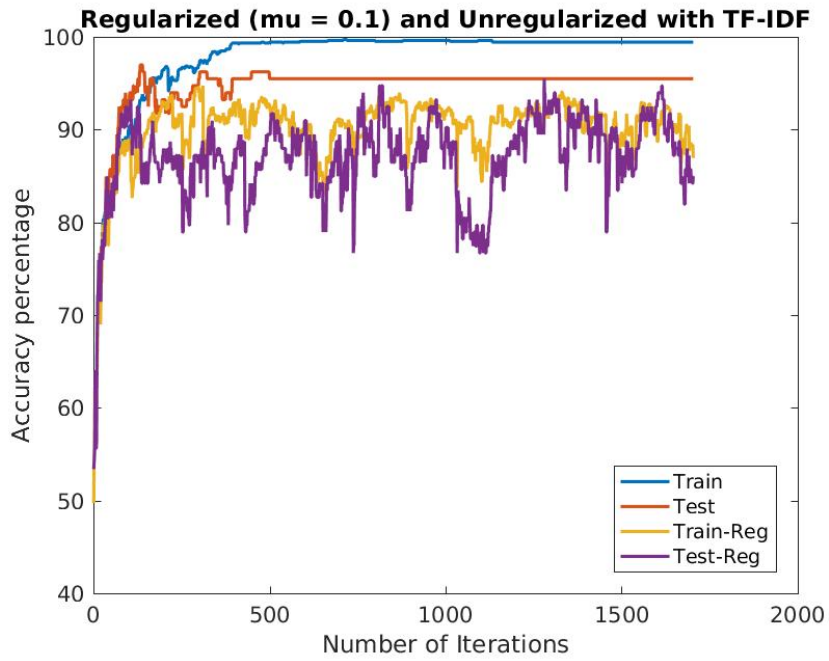
Answer 2: Without any extra credit options, the unregularized curve flattens out in **5 passes**.

Answer 3 and 4: Best predictors for Baseball class - **Baseball and Runs**. Best predictors for Hockey class - **Hockey and Playoffs**. Since this is binomial logistic regression, the best features for one class are the worst features for the other class

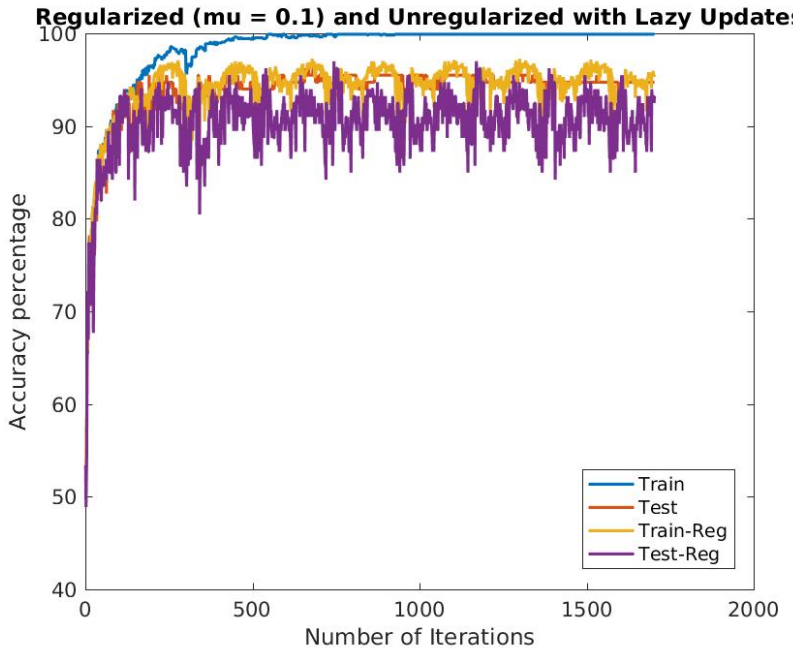
Extra Credit



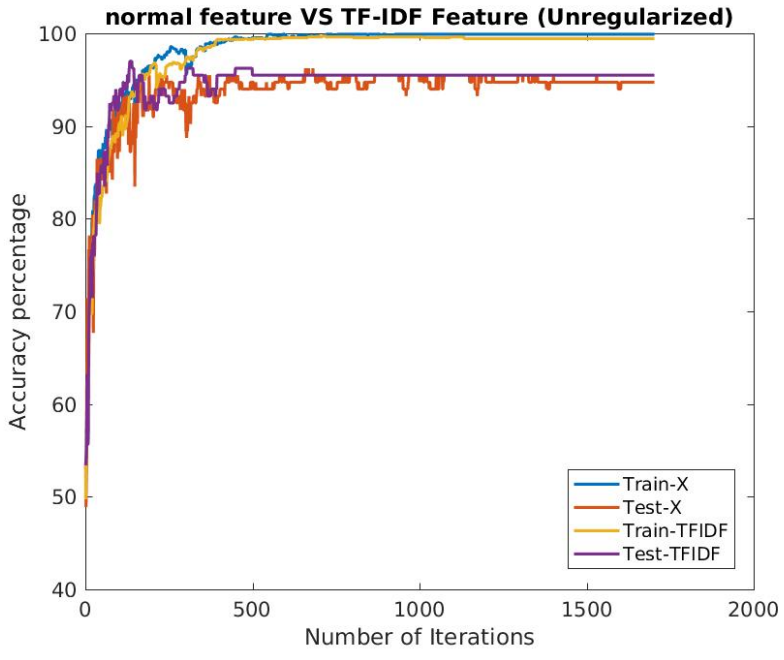
(a) Effect of Step Update



(b) Effect of TF-IDF



(c) Effect of Lazy Updates



(d) X vs TF-IDF. The Purple line(TF-IDF) indicates an accuracy of 95% and the red line(X) corresponds to an accuracy of 94%

Adaptive Learning Rate: $\frac{\text{args.step}}{\log(1+x)}$, where x is the iteration number. The idea is to decrease the step as it converges. The effect is seen in Fig (a).

Lazy Updates: Only modify the β values corresponding to non-zero feature values to decrease computational costs. The effect is seen in Fig (c)

TF-IDF: I use df = ft * ift where $\text{ft} = \frac{N}{\text{nt}} * \frac{1+\text{count}}{1+\text{total\_words\_in\_doc}}$ and $\text{ift} = \log(1+\frac{N-\text{nt}}{\text{nt}})$. N : Total number of documents, nt: number of docs in which term t appears