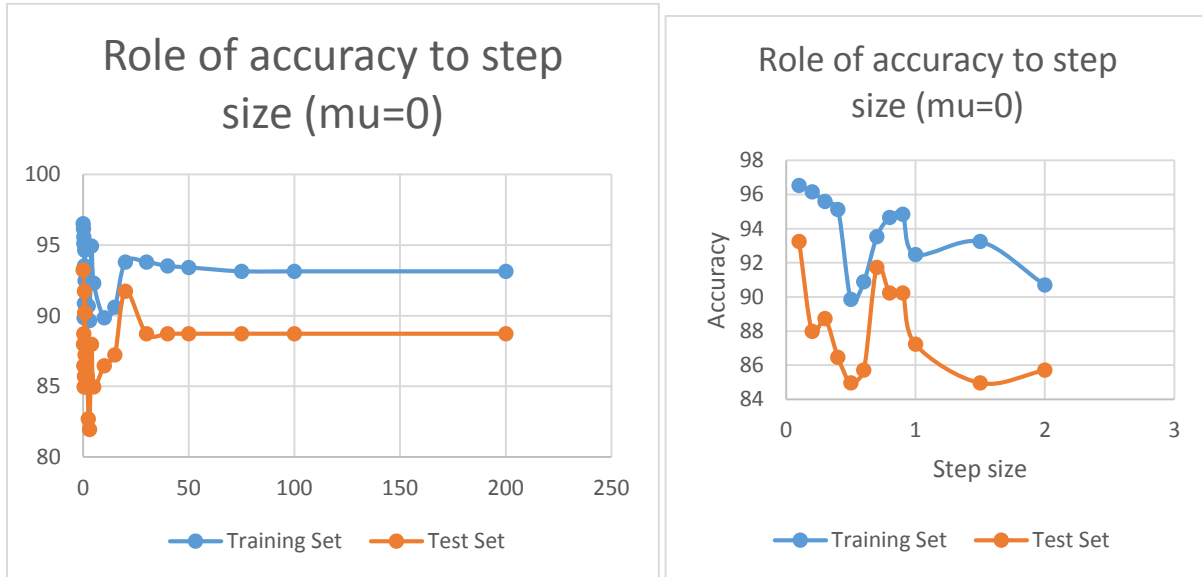
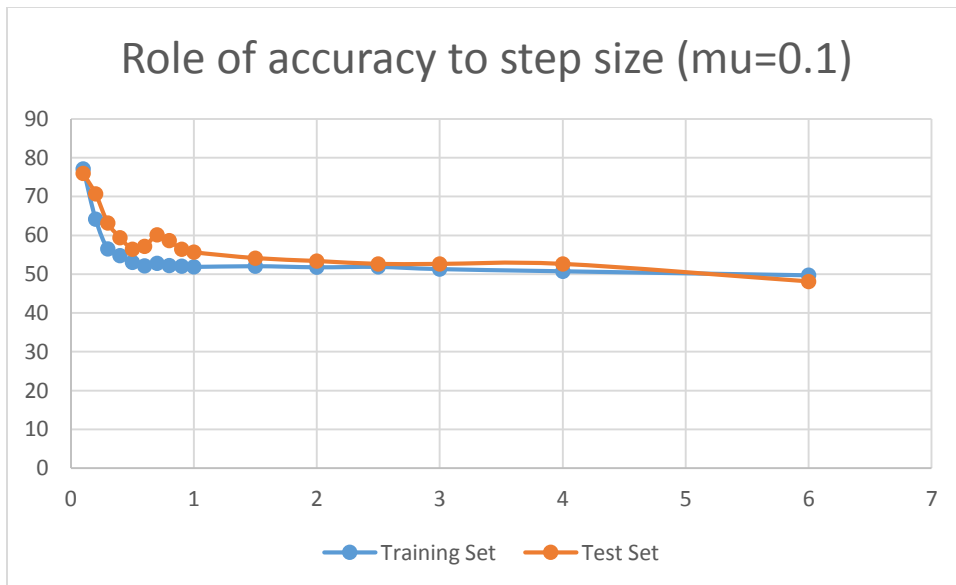


1. For this question as it is said in piazza I draw a graph for accuracy vs. step size:

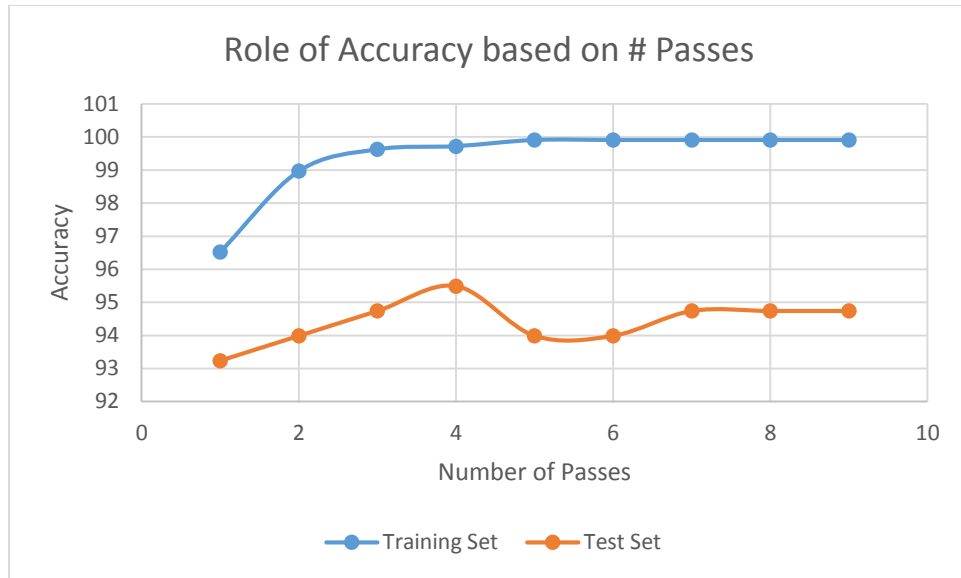
As you can see for unregulaized version by changing step size a little in the beginning the accuracy may increase or decrease. But when step size becomes equal to 75 and more the accuracy doesn't change anymore (this is because we consider $\exp(x) = \exp(700)$ if $x > 700$).



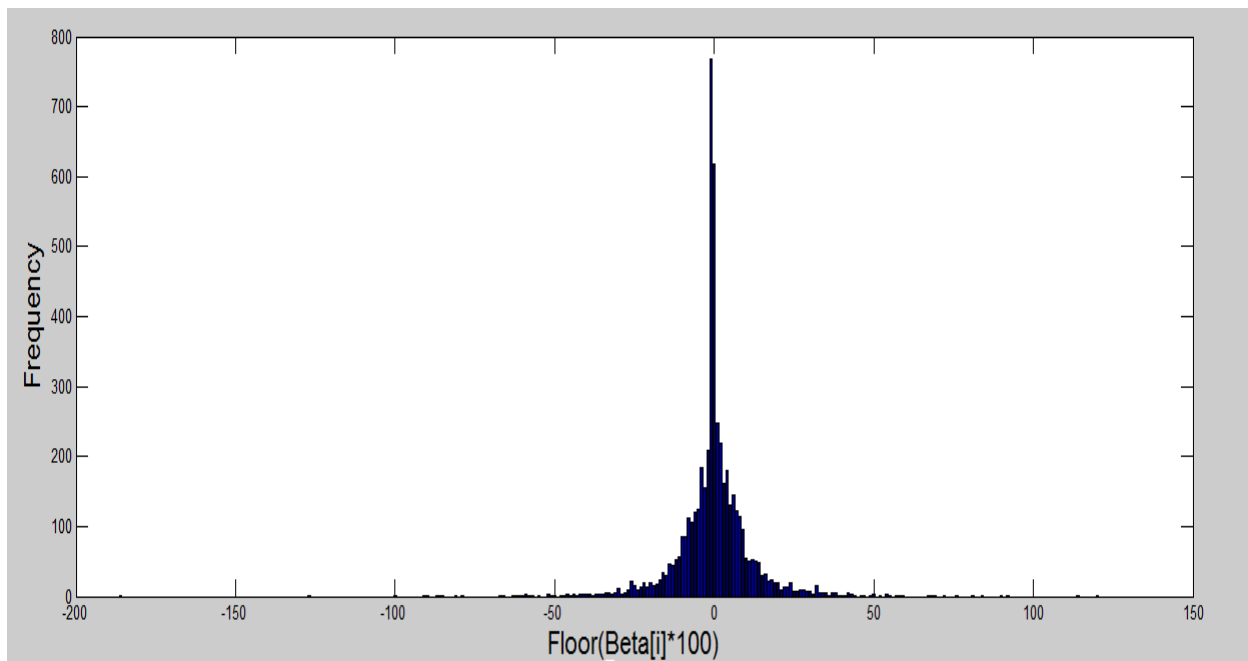
For regulized version, by increasing step size the accuracy will be decreased.



2. As you can see in the graph below, the test accuracy won't be increased after 5 passes and the test accuracy won't be increased after 7 passes. So we can tell that get highest possible accuracy for training set you need 5 passes over the data and to get highest possible accuracy for test set you need 7 passes.



3. Below you can see the histogram of beta. I calculate $\text{floor}(\text{beta}[i] * 100)$ first and then calculate the frequency of each of them. We know that if absolute value of $\text{beta}[i]$ is too large it is good predictor for positive class and if absolute value of $\text{beta}[i]$ is too large and negative it is a good predictor of negative class. From below graph we can see that the number of $\text{beta}[i]$ which its value is greater than 0.5 and less than -0.5 are very small. So the words who are related to these beta are best predictors:



Best predictors for negative class (those whose beta is less than -0.5)

**Hockey playoffs pick playoff points period pp golchowy goals next biggest roughing gld ice trades
never coverage coach contact cards broadcast round move mask goal captain head shown franchise
beat finals play**

Best predictors for positive class (those whose beta is greater than 0.5)

**hit runs bat saves pitching starters peak baseball ball age run usa hitting book strike still rather hitter
pitcher least talent uniforms bigs**

4.

The poorest predictors are those which their beta equals to zero:

**everywhere blasted intermissions bloody broad deceased hesitate hooked memoriam patterns pitiful racist
riel rode silence tone vintage wrestling**

After them, based on histogram of beta[i]s (the above figure) we can say that those beta which are positive and less than 0.001 are poorest predictors and those which are negative and their absolute values are less than 0.001 are poorest predictors.

Poorest predictors (Those whose beta is greater than 0 and less than 0.001) (we just mention some of them. There are 239 words whose beta is in this interval)

**tandem terrific partly dive kills chicago exaggerate reasonably essence surpass hurled slid chop
retiring benefits exceed**

Poorest predictors (Those whose beta is greater than -0.001 and less than 0) (we just mention some of them. There are 414 words whose beta is in this interval)

**kicked evetually possession smothered backhander diving dot booo misc barn clinched sleep spoiler
promoted profit shoveled trickling deflected skated neutralizing unscreened clouded enforcer killer
midway unit wednesday challenging singled valued pt stood.**