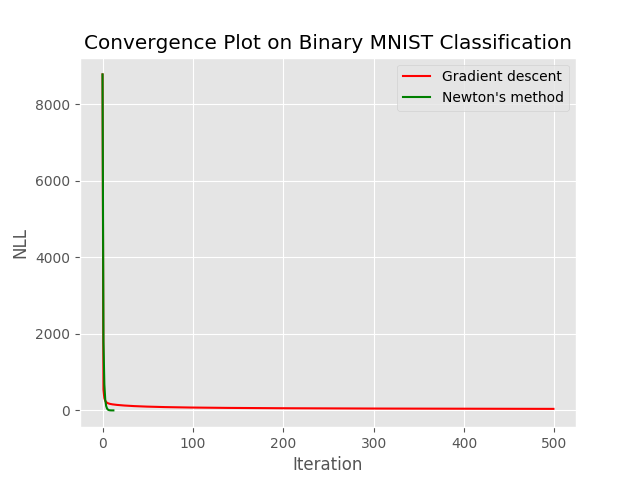
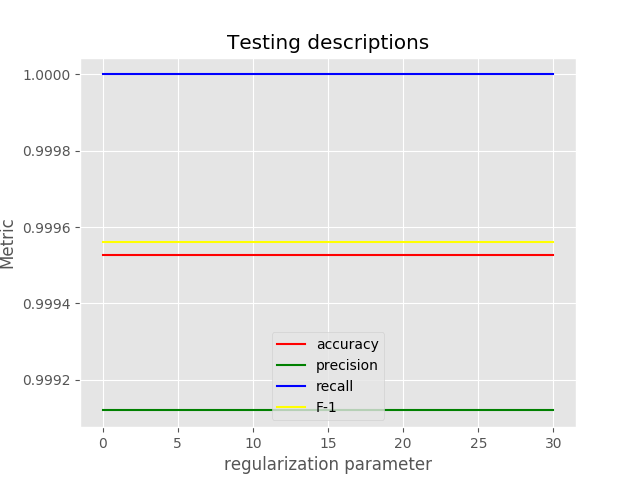
2. a)

Learning curve below



As we see, Newton’s method converges much faster than gradient descent!

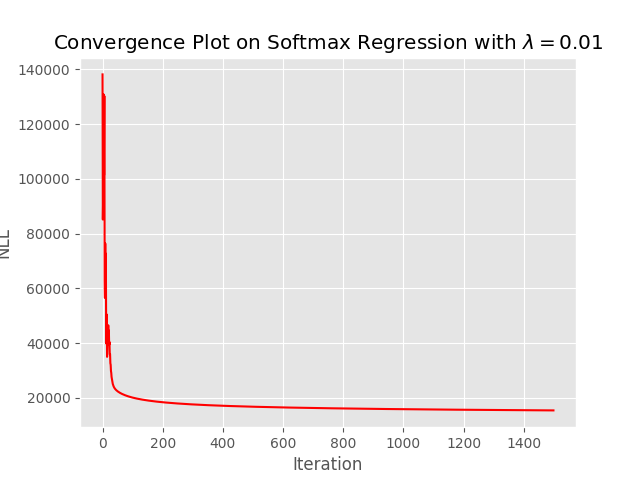
Metrics plot below



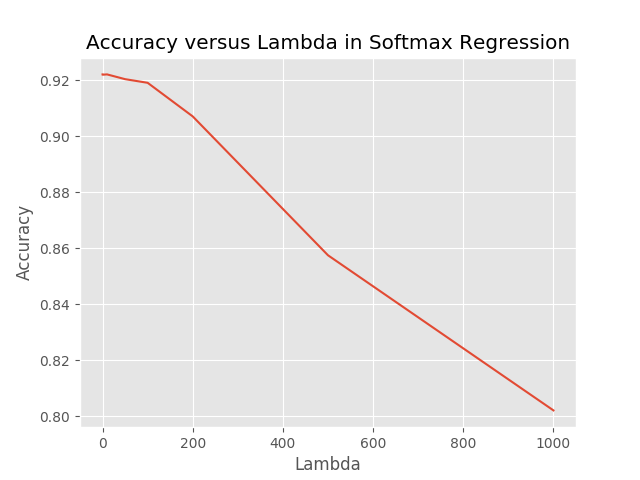
Accuracy on the test set is around 0.9995, which is amazing ☺

b)

Learning curve below



Accuracy plot below



We achieved an accuracy of 0.9221 on the test set with an optimal lambda of 0.01.