constant and the other parameters varied. From using the best hyper-parameters, where 10 epochs for each data set was chosen. The test set `astro/original/test` scored a classification accuracy of 0.90625, while the scaled data did much worse and `astro/scaled/test` had a classification accuracy of 0.50000. These results agree pretty well with the last homework assignment which is reassuring on the implemented algorithm working correctly. Figure 4 shows the negative log likelihood function of each of the test files at the end of each epoch.

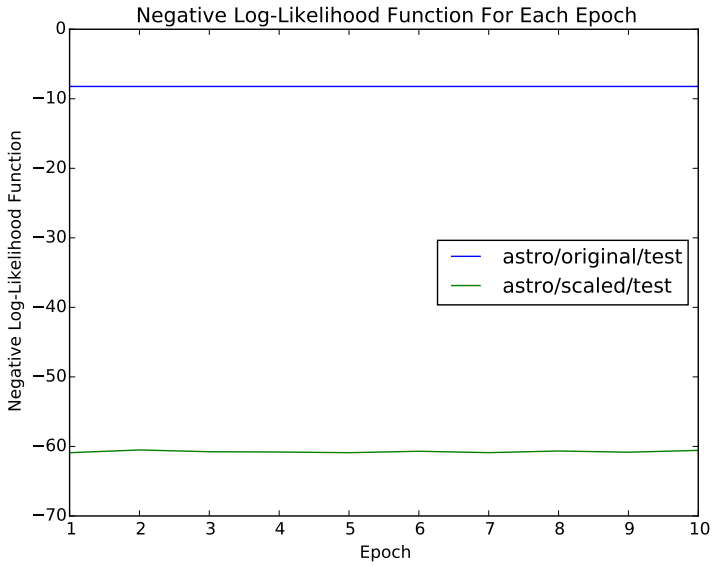

Figure 1: Negative Log Likelihood Function of $J(\mathbf{w})$

As mentioned in previous homeworks, you may use any programming language for your implementation. Upload your code along with a script so the TAs can run your solution in the CADE environment.

4 HAPPY HOLIDAYS Extra Credit

25pts You've seen stochastic gradient decent applied to logistic regression. Now we ask why this is a viable strategy for optimizing this objective. Prove that SGD will find the optimal value for this function. This can be done by demonstrating that the objective is convex. There are many ways to prove this. One of the most straightforward ways to show this is demonstrate that the Hessian is positive-semidefninite.

1. [5 points] Find the Gradient of:

$$\sum_{i=1}^{m} \log\left(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)\right) + \frac{1}{\sigma^2}\mathbf{w}^T\mathbf{w}$$