

1 Hyperparameter Testing

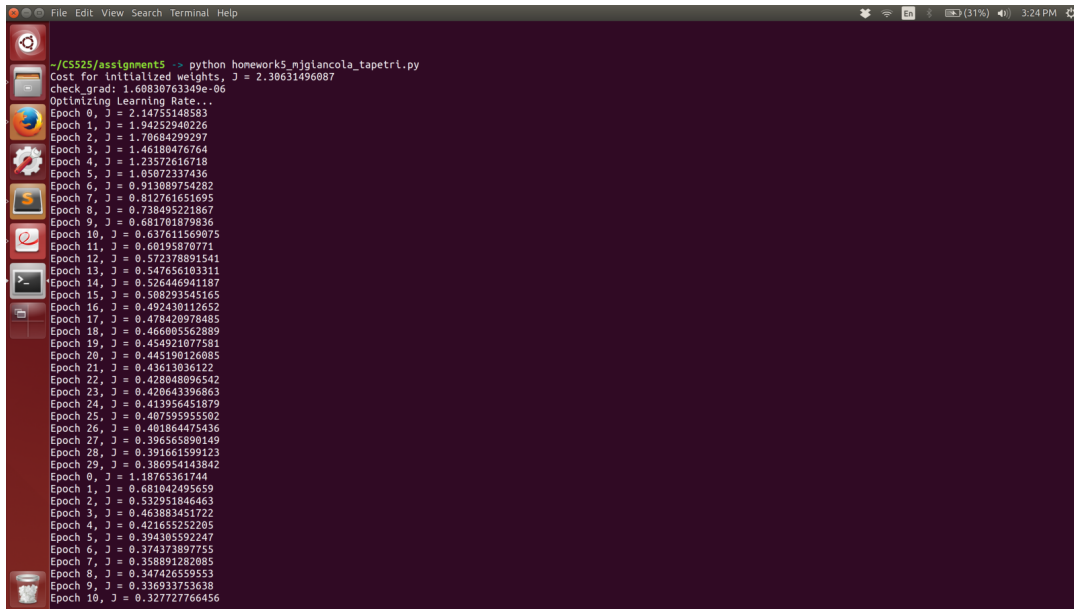
As there was a large number of possible hyperparameter settings, and we were intended to test at least 10 settings, we considered the following when designing our tests:

- Increasing the number of epochs generally improves performance
- Regularization didn't tend to improve performance (during casual testing of our network)

So, we chose to iterate over all of the suggested choices for number of hidden units, learning rate, and number of minibatches. We used 30 epochs during testing, as this seemed to be enough to get an idea for how well the other hyperparameters were performing, while not taking exhaustingly long. After testing, we chose 100 as the number of epochs for our model.

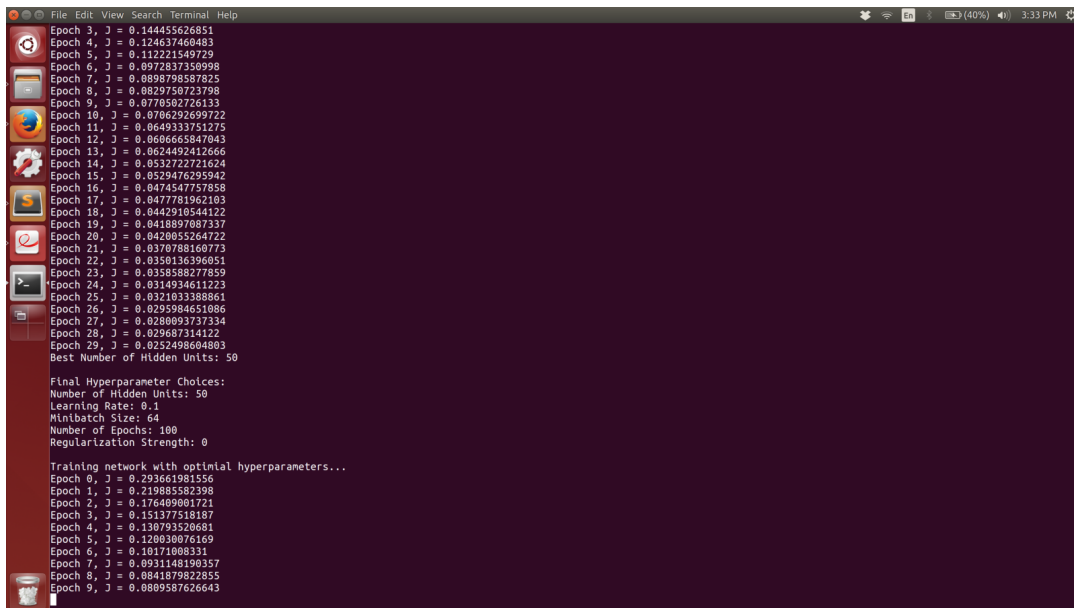
Clearly this approach has some flaws; namely, that the absolute "best" hyperparameter setting may be missed. However, testing every possible configuration would take a very long time, and this approach satisfied the requirements of the assignment.

2 Screenshots



```
~/CS525/assignments - python homework5_nigiancola_tapetri.py
Cost for initialized weights, J = 2.38631496887
check_grad: 1.60830763349e-06
Optimizing Learning Rate...
Epoch 0, J = 2.14755148583
Epoch 1, J = 1.24252940226
Epoch 2, J = 1.70684299297
Epoch 3, J = 1.46180476764
Epoch 4, J = 1.23572616718
Epoch 5, J = 1.05072337436
Epoch 6, J = 0.913089754282
Epoch 7, J = 0.812761651695
Epoch 8, J = 0.738495221867
Epoch 9, J = 0.681701879836
Epoch 10, J = 0.637611569075
Epoch 11, J = 0.60195870771
Epoch 12, J = 0.572378891541
Epoch 13, J = 0.547656183311
Epoch 14, J = 0.526446941187
Epoch 15, J = 0.508293545165
Epoch 16, J = 0.492438112652
Epoch 17, J = 0.478420978485
Epoch 18, J = 0.466085562889
Epoch 19, J = 0.454921077581
Epoch 20, J = 0.445190126085
Epoch 21, J = 0.43615036122
Epoch 22, J = 0.42804896542
Epoch 23, J = 0.420643396863
Epoch 24, J = 0.413956451879
Epoch 25, J = 0.407595955582
Epoch 26, J = 0.401864475436
Epoch 27, J = 0.396565890149
Epoch 28, J = 0.391661599123
Epoch 29, J = 0.386954143842
Epoch 0, J = 1.13765361144
Epoch 1, J = 0.681042495659
Epoch 2, J = 0.532951846463
Epoch 3, J = 0.463883451722
Epoch 4, J = 0.421655252285
Epoch 5, J = 0.394305592247
Epoch 6, J = 0.374373897755
Epoch 7, J = 0.358891282085
Epoch 8, J = 0.347426559553
Epoch 9, J = 0.336933753638
Epoch 10, J = 0.327727766456
```

Figure 1: Starting Hyperparameter Optimization



```
Epoch 3, J = 0.144455626851
Epoch 4, J = 0.124637460483
Epoch 5, J = 0.112221549729
Epoch 6, J = 0.0972837350598
Epoch 7, J = 0.0898798587825
Epoch 8, J = 0.0829750723798
Epoch 9, J = 0.0770582726133
Epoch 10, J = 0.0706292699722
Epoch 11, J = 0.0649333751275
Epoch 12, J = 0.0606665847043
Epoch 13, J = 0.0624492412666
Epoch 14, J = 0.0532722721624
Epoch 15, J = 0.0529476295942
Epoch 16, J = 0.0474547757858
Epoch 17, J = 0.0477781962103
Epoch 18, J = 0.0442910544122
Epoch 19, J = 0.0418897087337
Epoch 20, J = 0.0420855264722
Epoch 21, J = 0.0370788168773
Epoch 22, J = 0.0358136396851
Epoch 23, J = 0.0358588277859
Epoch 24, J = 0.0314934611223
Epoch 25, J = 0.0321033388861
Epoch 26, J = 0.0295984651886
Epoch 27, J = 0.0280093737334
Epoch 28, J = 0.029687314122
Epoch 29, J = 0.0252498604803
Best Number of Hidden Units: 50

Final Hyperparameter Choices:
Number of Hidden Units: 50
Learning Rate: 0.1
Minibatch Size: 64
Number of Epochs: 100
Regularization Strength: 0

Training network with optimal hyperparameters...
Epoch 0, J = 0.293661981556
Epoch 1, J = 0.219885582398
Epoch 2, J = 0.176409801721
Epoch 3, J = 0.151377518187
Epoch 4, J = 0.130793520681
Epoch 5, J = 0.120030076169
Epoch 6, J = 0.10171080331
Epoch 7, J = 0.0931148190357
Epoch 8, J = 0.0841879822855
Epoch 9, J = 0.0809587626643
```

Figure 2: Finished Hyperparameter Optimization



Figure 3: Results on Test Set