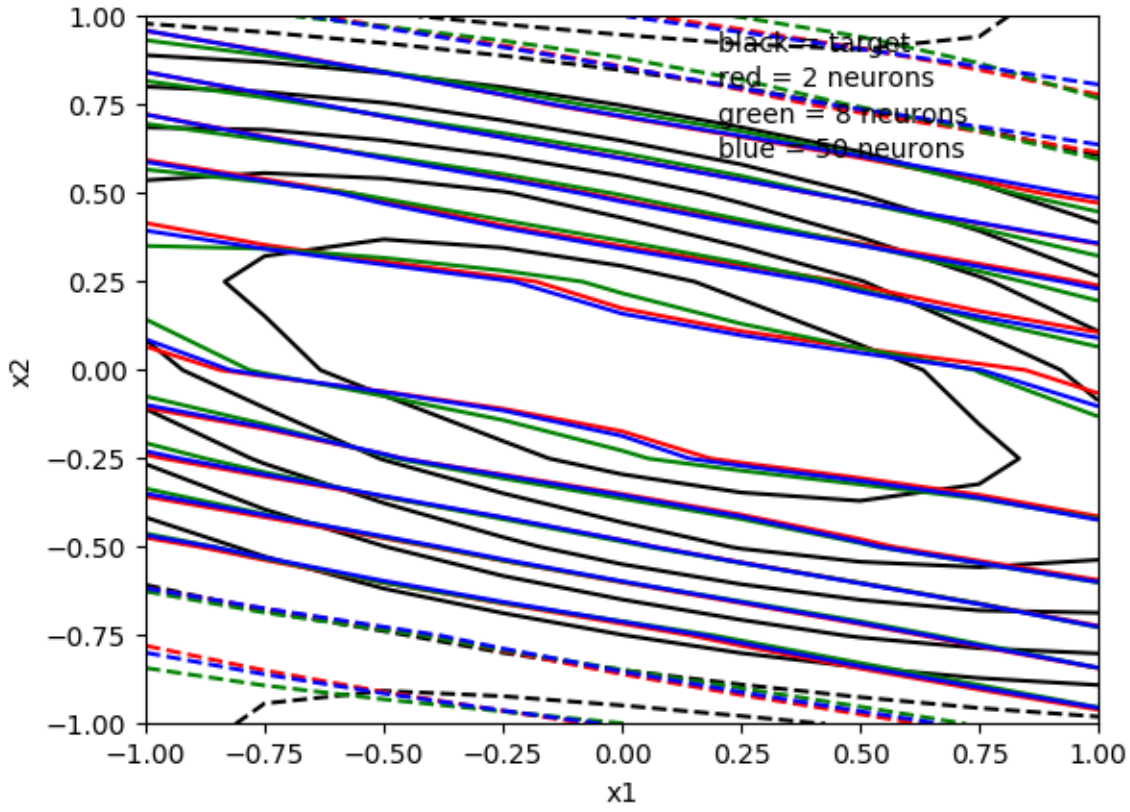
Question 1a

1. During my testing, I found that the 8-neuron model converged consistently and more rapidly than the 2-neuron and 50-neuron variants. Results were obtained over 5 iterations of 1000 epochs.

2.



3.

```
(base) D:\school\4107\a2>python q1.py
2 neurons converged after 582.4 epochs on average, at an average MSE of 0.0524793466553092 .
8 neurons converged after 270.4 epochs on average, at an average MSE of 0.019804502185434104 .
50 neurons converged after 520.2 epochs on average, at an average MSE of 0.019720953702926636 .
```

Note that the 2-neuron variant has an average MSE higher than the threshold for convergence, meaning that it must have failed to converge on at least one occasion.

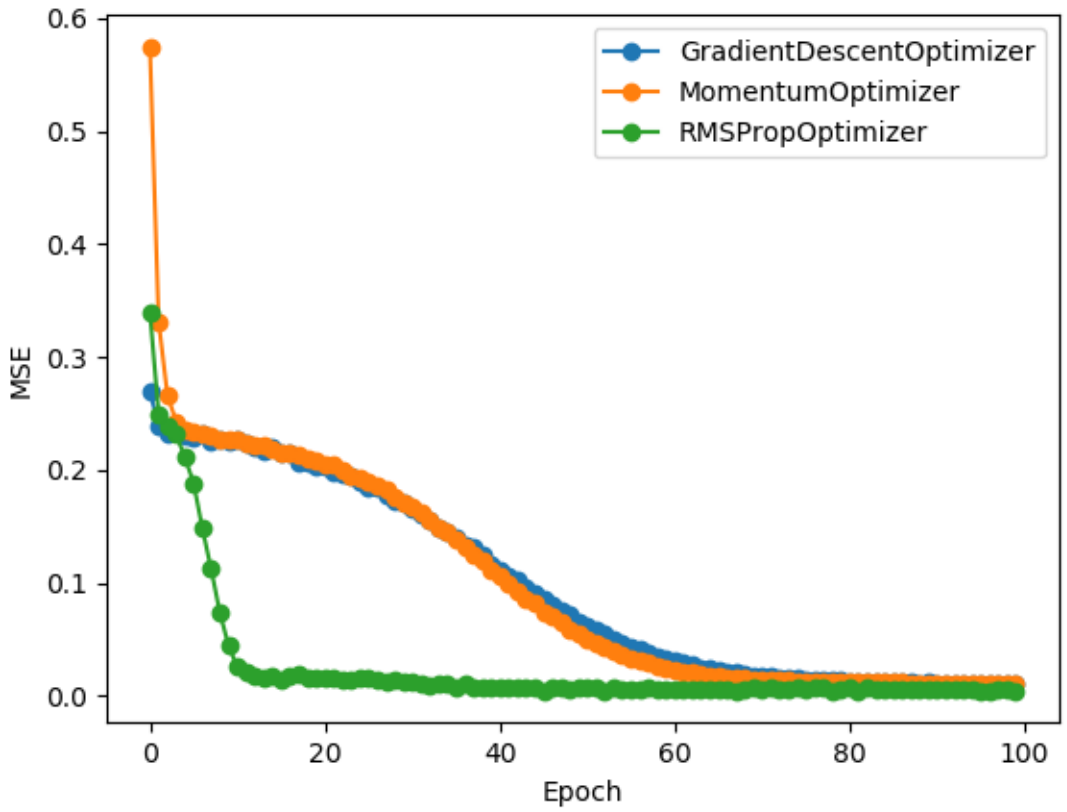
4. Results were obtained using the hyperbolic tangent activation function.

Question 1b

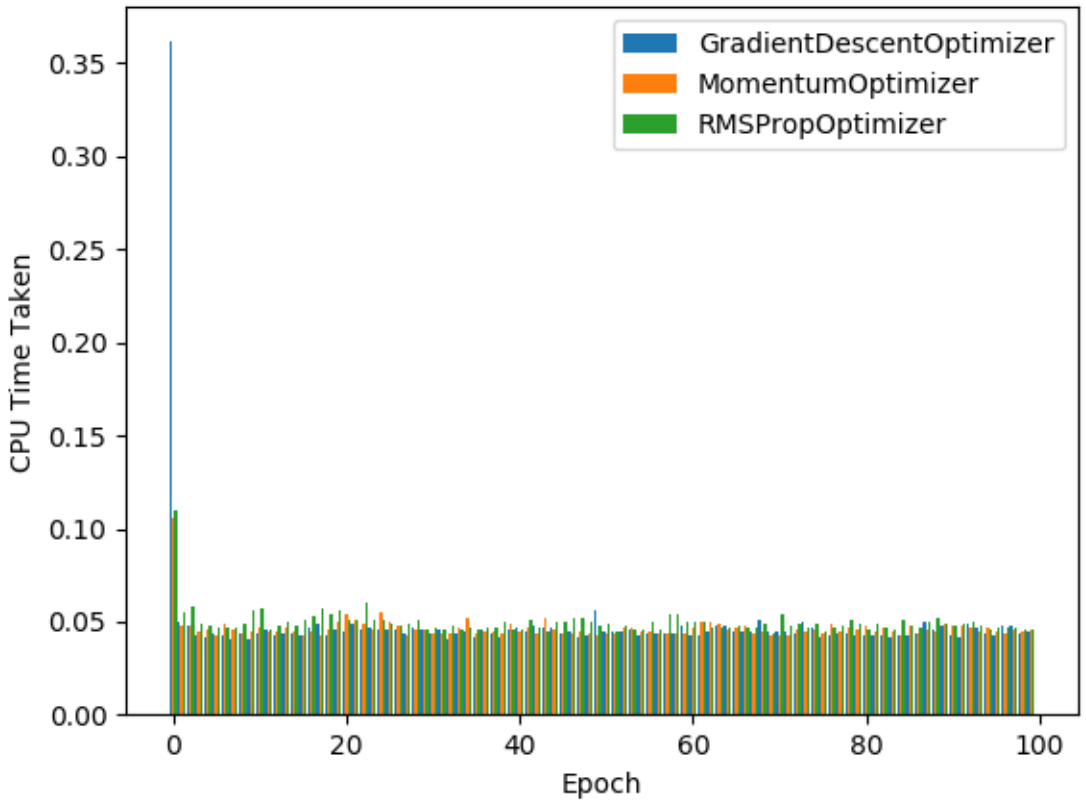
1. Note that per the specification, these results were obtained off of one 100-epoch training session, and may fluctuate depending on the initial values of the weights when you run q1.py yourself. Also, running b() in q1.py will generate 300 .hdf5 checkpoint files in the current directory so that the models can revert to when they initially converged; you have been warned.

```
Using TensorFlow backend.
Gradient descent took 69 epochs to converge.
The momentum optimizer took 64 epochs to converge.
RMS took 13 epochs to converge.
```

2. Graph obtained from same session as above. RMS clearly converges much earlier than the other 2, which are virtually tied.



3. Each optimizer is slow to start up (relatively speaking), and GradientDescentOptimizer takes the longest by far; however, they are all within approximately 0.05 seconds of each other during subsequent epochs.



4. Due to the random nature of training, no one classifier is 100% likely to have the greatest accuracy post-training. However, RMS is consistently superior over multiple trials. It scored particularly high on the session in question. The momentum optimizer is much less likely to be the best, and gradient descent less likely than that.

```
AFTER 100 EPOCHS
Gradient descent achieved RMSE of 0.07924511 on the test set.
The momentum optimizer achieved RMSE of 0.0858362 on the test set.
RMS achieved RMSE of 0.038231634 on the test set.
RMS had the best accuracy at the end of training.
```

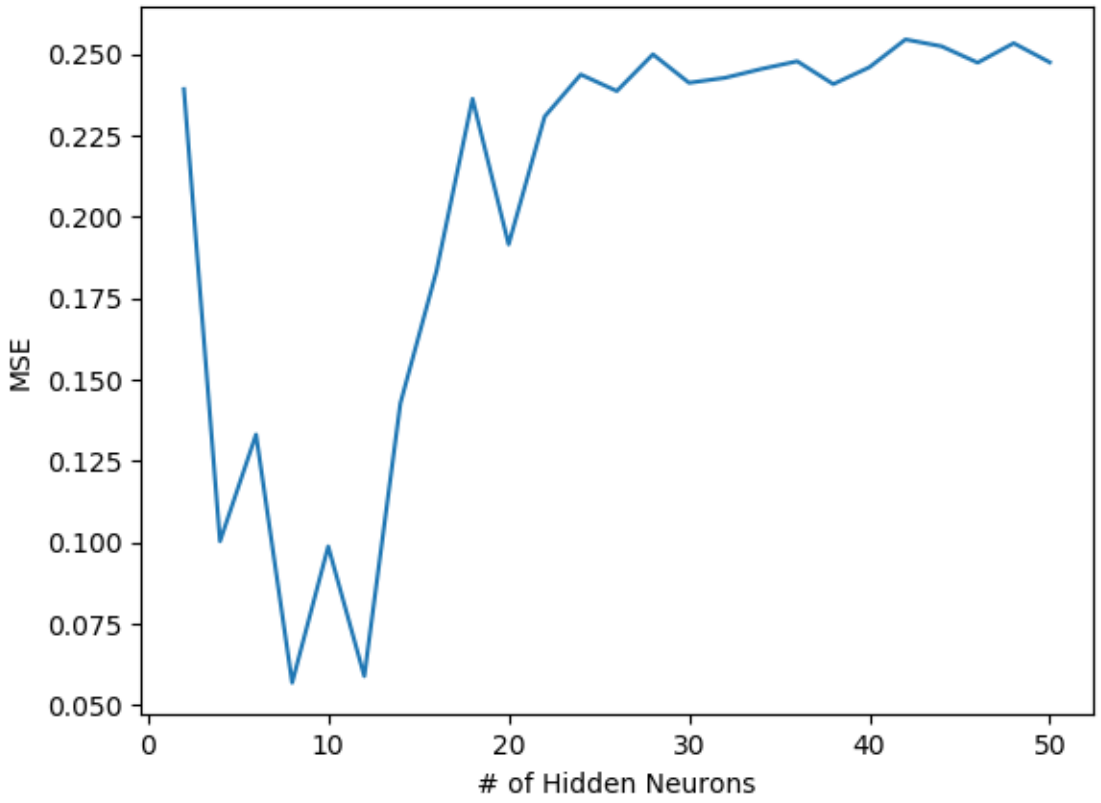
5. Similar to above, RMS is also typically the best option when tested at the point of convergence, meaning it is not only the most accurate, but the quickest to train.

```
AFTER CONVERGENCE
Gradient descent achieved RMSE of 0.112471804 on the test set.
The momentum optimizer achieved RMSE of 0.11286559 on the test set.
RMS achieved RMSE of 0.10001596 on the test set.
RMS had the best accuracy at the time of convergence.
```

Question 1c

1. MSE, as well as computation time, is optimal at between 8-12 neurons; anything past that is more expensive and less accurate.

2. Results obtained over an average of 5 250-epoch sessions on 25 different models, from 2 to 50 neurons.



3.

Fig. 6 reproduction

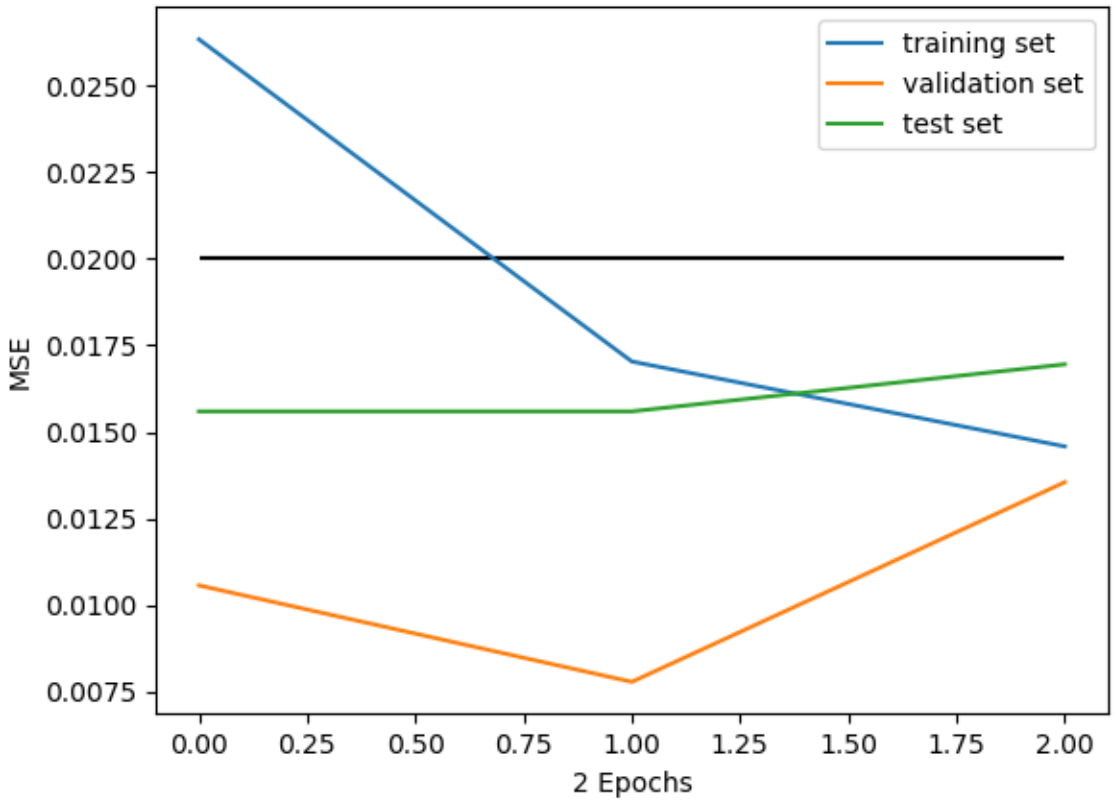


Fig. 7 reproduction