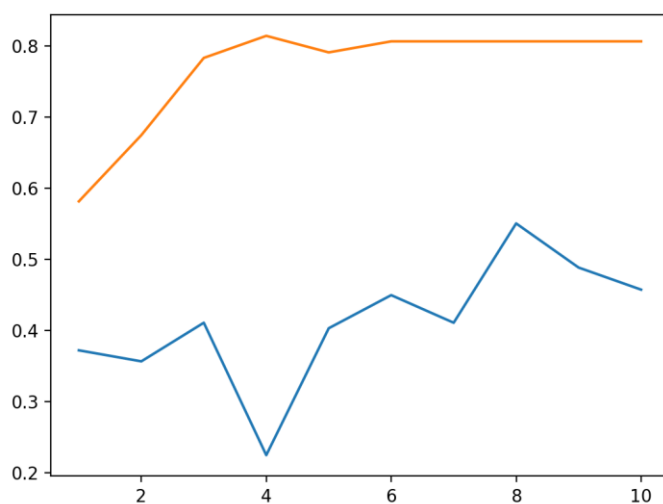


#### Question 4

The hidden layer sizes are 200 on the first hidden layer and 50 on the second hidden layer  
We chose to use the first 75 components of PCA since more PCA is not as much needed and 75 components is more than enough to show the superiority of PCA in training speed. Changing the components to other values will still allow it to work.

```
input layer size: 1850 h1 size: 200 h2 size: 50 output size: 7
raw data trained
2018-11-25 19:58:43.531167: I tensorflow/core/platform/cpu_feature
epoch: 0 accuracy: 0.37209302325581395
epoch: 1 accuracy: 0.35658914728682173
epoch: 2 accuracy: 0.4108527131782946
epoch: 3 accuracy: 0.2248062015503876
epoch: 4 accuracy: 0.40310077519379844
epoch: 5 accuracy: 0.4496124031007752
epoch: 6 accuracy: 0.4108527131782946
epoch: 7 accuracy: 0.5503875968992248
epoch: 8 accuracy: 0.4883720930232558
epoch: 9 accuracy: 0.4573643410852713
input layer size: 75 h1 size: 200 h2 size: 50 output size: 7
optimized pca trained
epoch: 0 accuracy: 0.5813953488372093
epoch: 1 accuracy: 0.6744186046511628
epoch: 2 accuracy: 0.7829457364341085
epoch: 3 accuracy: 0.813953488372093
epoch: 4 accuracy: 0.7906976744186046
epoch: 5 accuracy: 0.8062015503875969
epoch: 6 accuracy: 0.8062015503875969
epoch: 7 accuracy: 0.8062015503875969
epoch: 8 accuracy: 0.8062015503875969
epoch: 9 accuracy: 0.8062015503875969
```

Accuracy:



The Orange is PCA optimized

The Blue is Raw data

As you can see the orange data consistently outperforms raw data from the start.