

Assignment 1 Part 1c)

Images per class in the sample of 12000 images with seed=5:

```
#Summary of number of images per class
from collections import Counter
Counter(y_test[:,0])

Counter({0: 1207,
        1: 1177,
        2: 1184,
        3: 1158,
        4: 1223,
        5: 1202,
        6: 1206,
        7: 1219,
        8: 1180,
        9: 1244})
```

Performance of the best model created on the new data set of 12000 images :90%

Accuracy of the model trained using 12000 images: 63.46%

The accuracy of the model is higher in the first case, because of the randomness in the sampling in second case. This might result in bias-variance tradeoff as we are selecting random numbers that are optimized for a small sample of data. The quality of such models is dependent on the quality of the dataset that has been held out. In the production, we need to select a model that has been trained on the entire training dataset, once we decide an algorithm and configuration.