Question 2 write up

Amanda Shen, Eunjun Choo, Michelle Liu

2 (i)

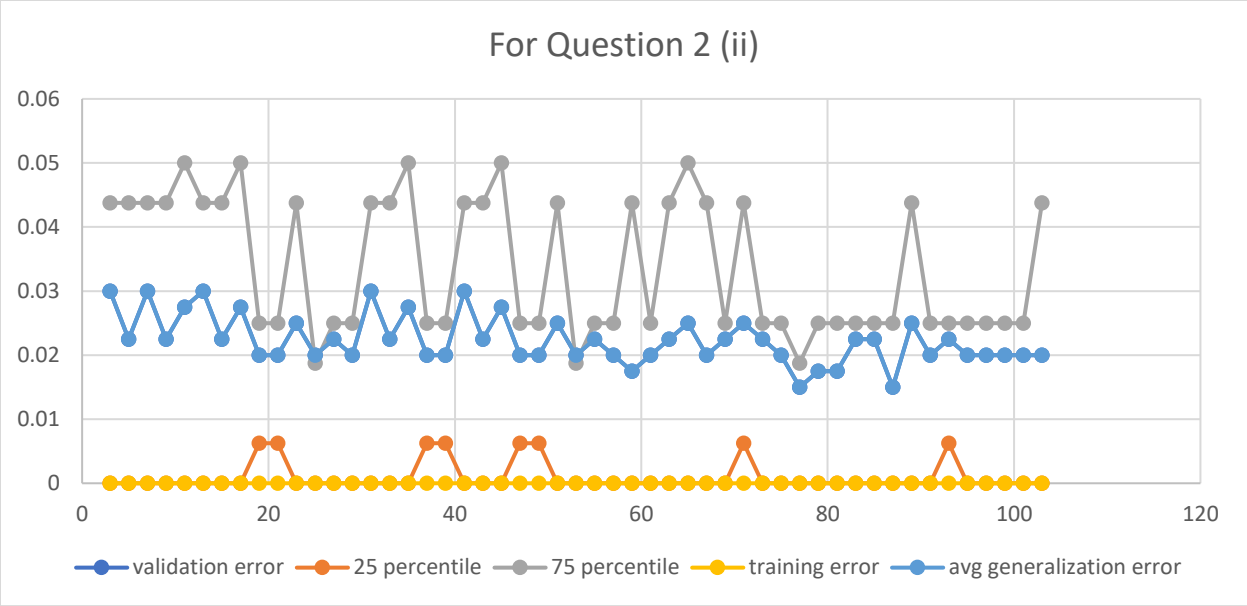
To check, we have our pictures as follows. The check is under “data_setup” function and the commented-out portion that indicates that the chunk of code is for checking the picture

```
# for test 9
nump_te_9 = te_9.to_numpy()
mat_te_9 = nump_te_9[1].reshape(28, 28)

## Check for the picture.... For 2 (i)
""" Image.fromarray(np.uint8(mat_tr_5), "L").save("trial5", "JPEG")
Image.fromarray(np.uint8(mat_tr_9), "L").save("trial9", "JPEG")

Image.fromarray(np.uint8(mat_te_5), "L").save("test5", "JPEG")
Image.fromarray(np.uint8(mat_te_9), "L").save("test9", "JPEG")
"""
```

We see that our formatting is correct.

2 (ii)

After applying our KNN-classifier and cross-validation, we plot the training and CV misclassification rate for odd values of K from 3 to 103



We found that for k = 77, the validation error/avg generalization error was the lowest.

2 (iii)

With our chosen k as 77, our result is as follows:

Each fold performance	Average Training Error	Generalization error
0.02577319587628866	0	0.02577319587628866