HW 11 Written Work

● Graded

Student

Sangwon Ji

Total Points

20 / 20 pts

Question 1

Q1.7 Resolved **10** / 10 pts

✔ **+ 3 pts** Training set is for building the classifier.

✔ **+ 3 pts** Test set is to measure performance of the model on unseen data / how well the model generalizes.

✔ **+ 4 pts** We shouldn't use the test set more than once since it might bias our model, causing the model to overfit / not generalize well to unseen data/is unnecessary because we already know the accuracy.

+ 0 pts Blank/Incorrect

Regrade Request **Submitted on: Aug 16**

Hi,
I think I have mentioned that the training set is for building the classifier as I have mentioned that it works differently as to see the difference in the real world. That is to do with building the classifier to test and I think it makes sense without the exact sentence..
Thank you!

Fixed

Reviewed on: Aug 16

Question 2

Q1.8 **10** / 10 pts

✔ **+ 10 pts** Correct. Mention it avoids ties or that it guarantees a majority.

+ 0 pts Blank/Incorrect.

Question 1.7. Why do we divide our data into a training and test set? What is the point of a test set, and why do we only want to use the test set once? Explain your answer in 3 sentences or less. **(10 points)**

Hint: Check out this section in the textbook.

The reason we divide our data into a training and test set is because we want our algorithm to be able to generalize, and having the test set is testing whether our algorithm can do that. The role and the point of a test set, is a way to understand how well our classifier would perform in a real world scenario with the data it has not seen before. We only want to use the test set once since it represent the data that is out of sample, that provides the best possible estimate, that can be used obejctively.

Question 1.8. Why do we choose `k` to be an odd number in k-NN? Explain. **(10 points)**

If we choose an even number, there is chance that there might be ties when evaluating the data. We will not be able to classify if the results come out tie. For example having 6 for k, we might get the same results for 3 and 3 for both standford and berkeley.