

**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.

*Type your answer here, replacing this text.*

**SOLUTION:** We divide our data into a training and test set, so we can use the training set in order to build our classifier, and the test-set is used to see if your model will be able to generalize to data you have never seen before. Ideally, you should only use the test set once; otherwise, you are biasing your classifier and it may learn how to get a good accuracy on your training/test set, but not generalize to data you have never seen before. (You'll learn more about this in depth if you decide to take Data 100)



**Question 1.8.** Why do we use an odd-numbered  $k$  in  $k$ -NN? Explain. (10 points)

*Type your answer here, replacing this text.*

**SOLUTION:** We use an odd-numbered  $k$  in  $k$ -NN in order to avoid having ties; using an odd-numbered  $k$  guarantees a majority.

