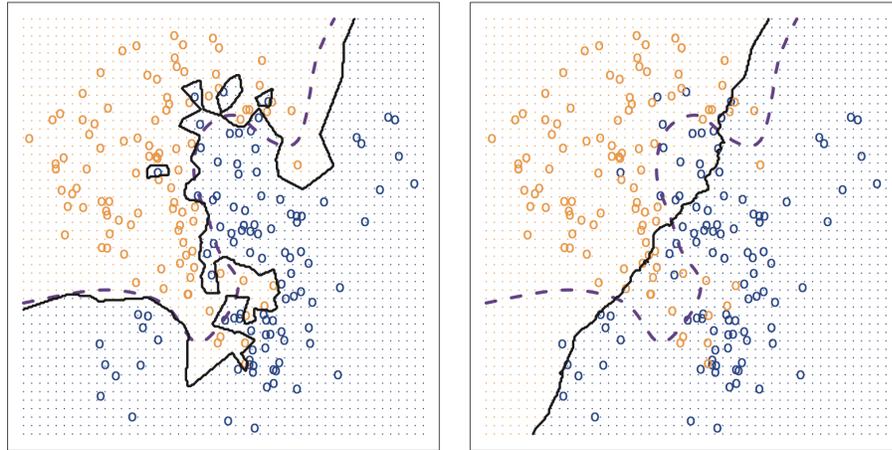


**K-nearest neighbors and KD-trees**

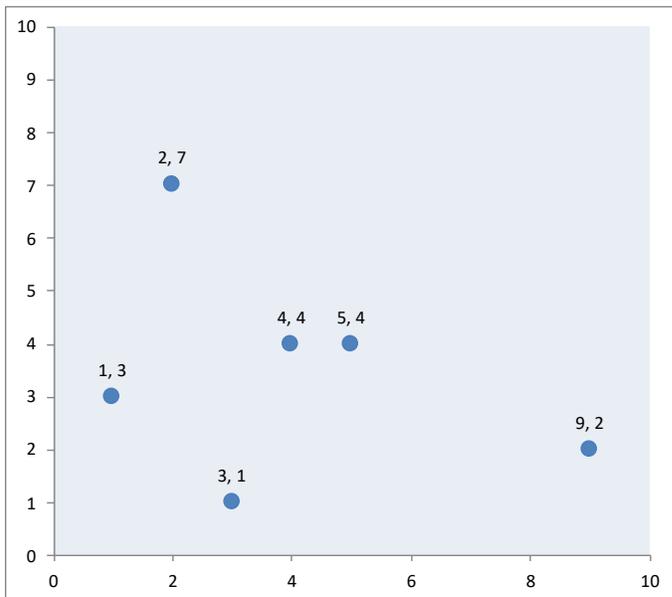
(find and work with a partner)

- In the two images below, the solid lines represent the KNN *decision boundaries* (the dashed lines represent “ideal” decision boundaries – we will discuss this later). In each case, what is your guess for the approximate value of  $K$  that was used?



- In the scenarios above, which would you describe as *underfitting*? As *overfitting*?

- For the given training point below in 2D, create a KD-tree, starting with the x-axis.



- For the test points  $(7,7)$  and  $(5,0)$ , show how the KD-tree would be used to find the nearest neighbor.

5. What is the runtime of finding the nearest neighbor for a single test point, if there are  $n$  training points and  $d$  total dimensions? Answer for:

(a) The naive KNN algorithm.

(b) Using a KD-tree.