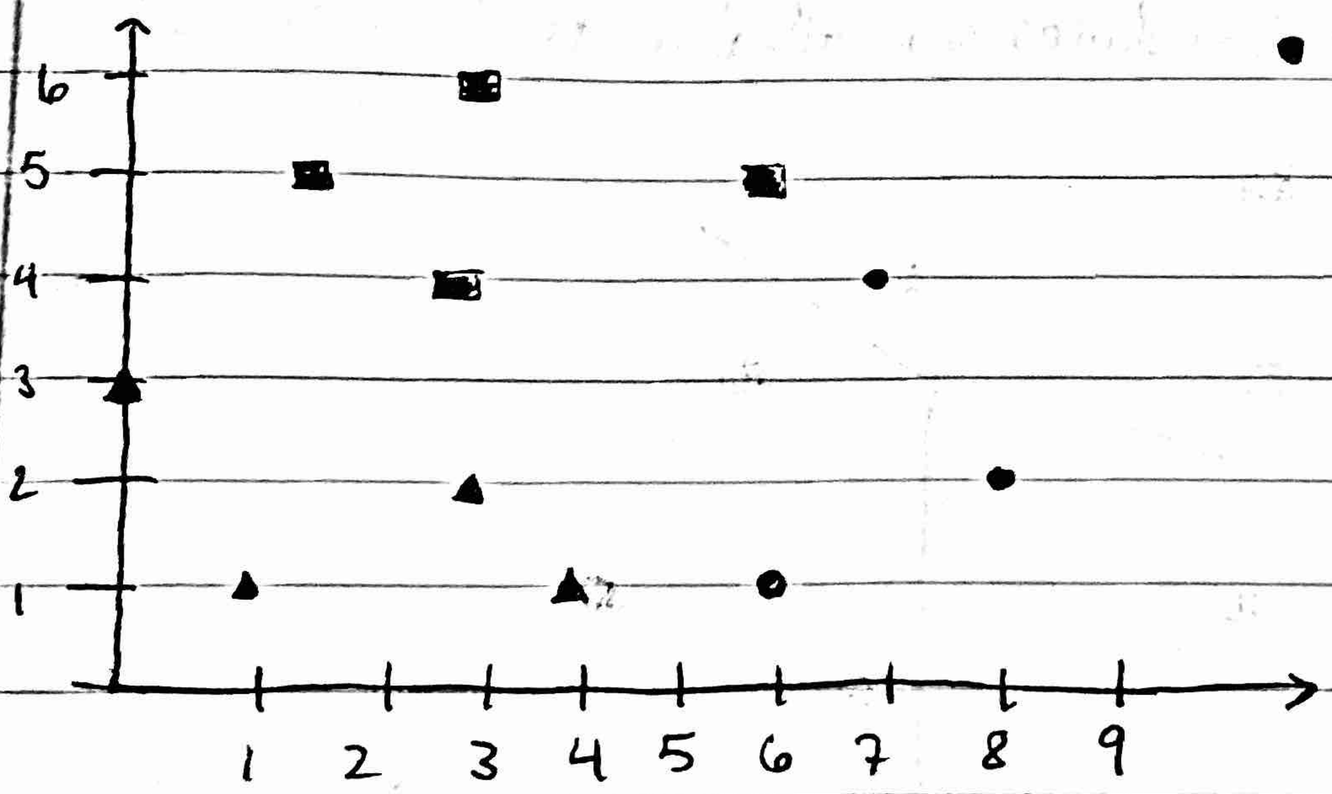
K-Nearest Neighbors Discussion #4

Goal of KNN

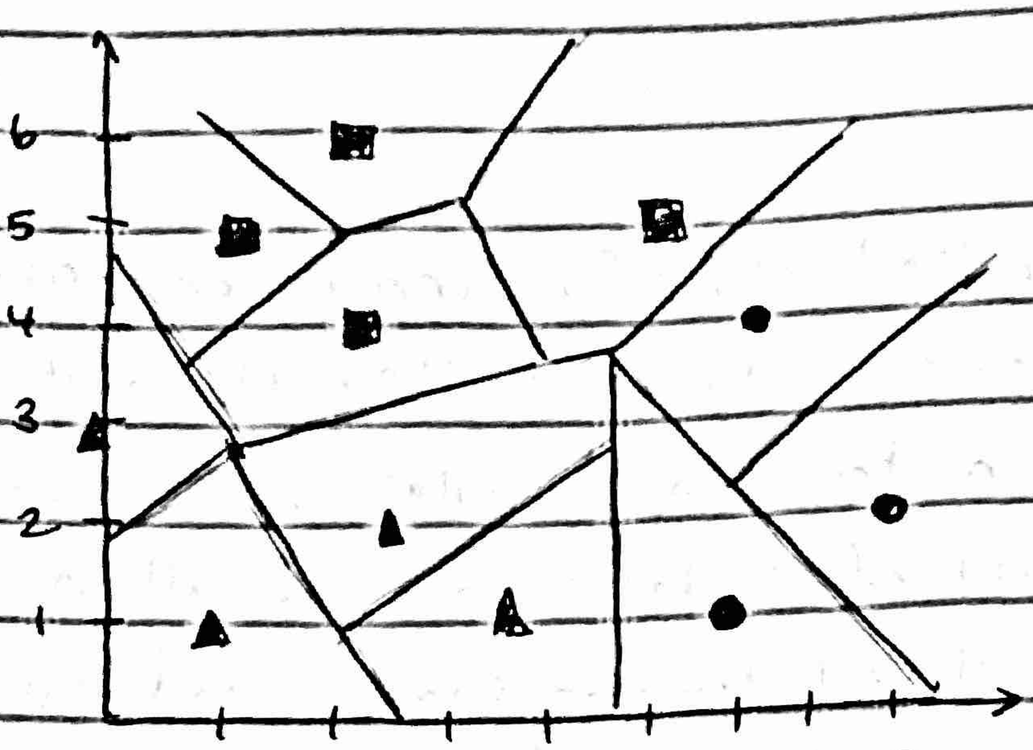
KNN is a supervised classification algorithm. (Using the labels of the training data, it attempts to classify a test point). Given n training samples, we compute the distance (usually Euclidean) of the test point to all n training samples. We pick the K closest training samples and classify the test point via majority vote.

Example



Q: Draw the decision boundary. How?

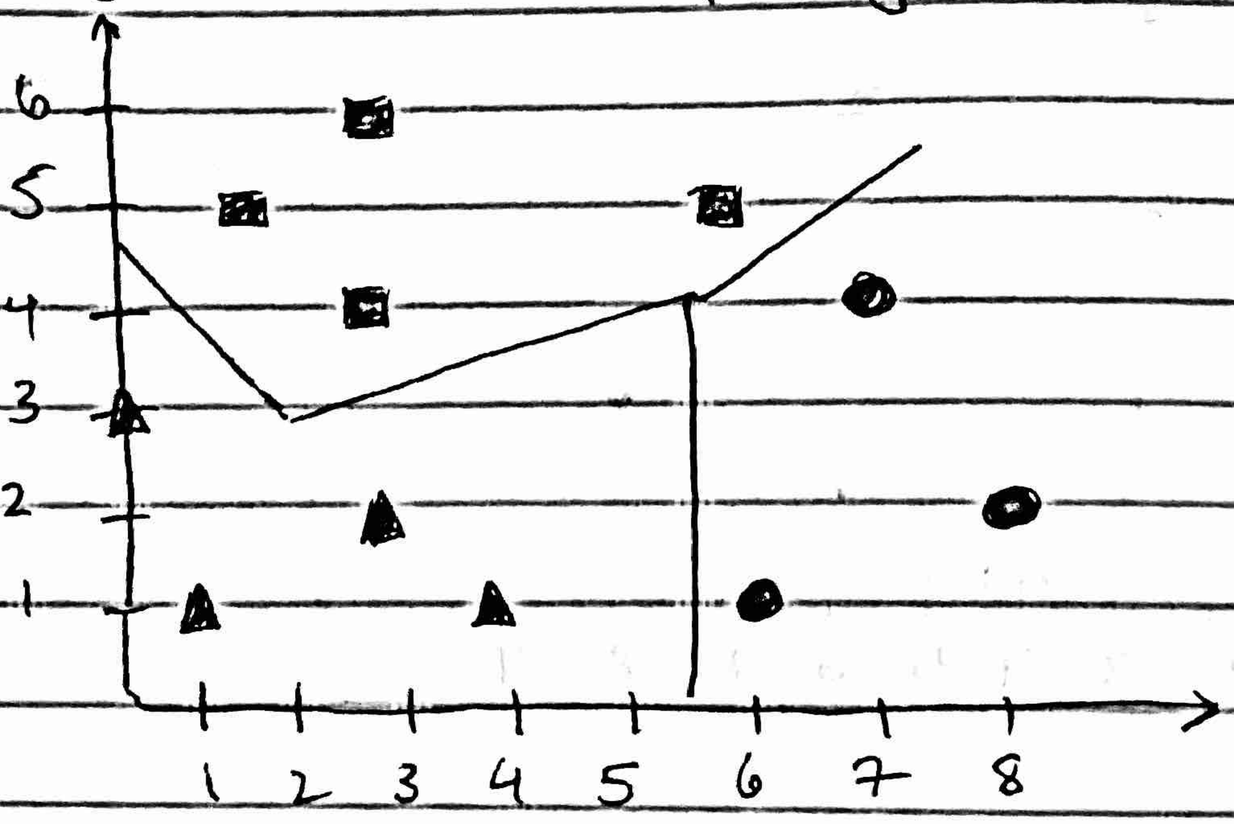
1. Draw perpendicular bisectors between each pair of neighboring points. The resulting regions are called Voronoi cells.



(Not an exact diagram)

Should find the points of intersection for each (this will make it easier to classify test points)

2. Consolidate regions belonging to the same class
(Remove boundaries separating points of same class)



Q: classify a test point. How?
Plot the test point and classify based on which region it is in.
(For a 1-NN classifier, the closest neighbor will be in the class region of the test point.)

What if I don't want to plot / can't plot my test points?

1. Determine k
2. Calculate distance between test point and all training samples (Euclidean)
3. Determine k closest training points
4. Choose whichever class is majority among k closest neighbors as the test point's class.

No majority / on decision boundary? Pick arbitrarily!