# K-Nearest Neighbors

We use the k nearest points, either equally weighted or weighted by distance or things like this. This is a way of using existing data and saying things that have the most similar features.

We may want to experiment with different K values by reserving some training data for testing and validation

K too low makes the data easy to influence by outliers

K too large is very smoothing, also if the data is not balanced it is way more likely that classes with more points are going to be picked as they will correspond to more points, even if the data is way closer to the cluster of a different class.

# Hierarchical Clustering