

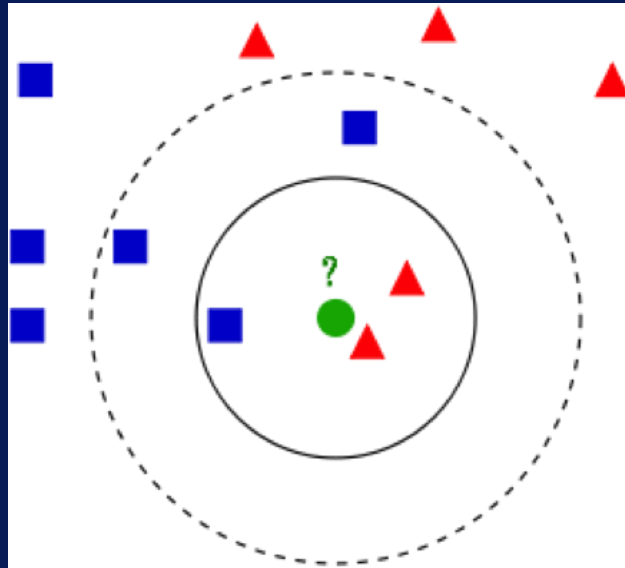
k Nearest Neighbors

After this video you will be able to..

- Describe how kNN is used for classification
- Discuss the assumption behind kNN
- Explain what the 'k' stands for in kNN

kNN

- Simple classification technique
- Label sample based on its neighbors



kNN Assumption

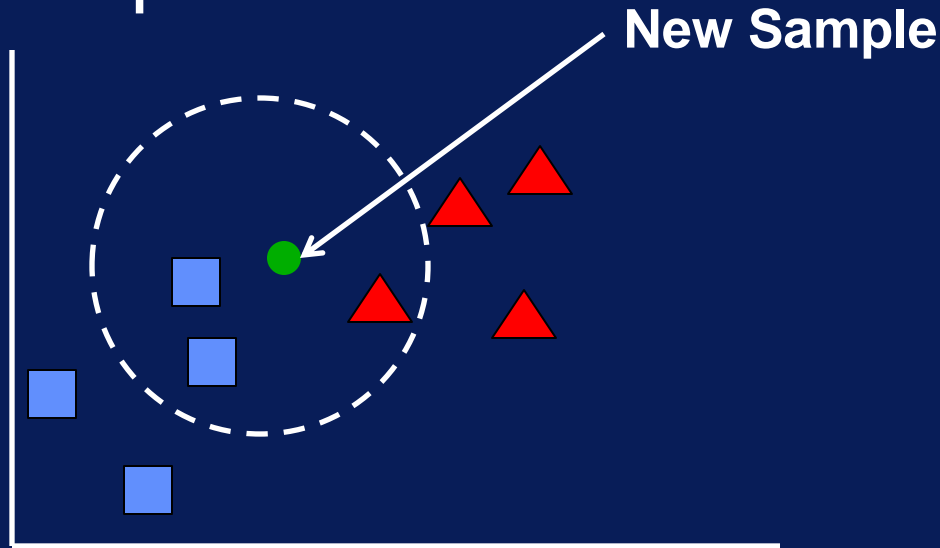
- Duck test

Quack



How kNN Works

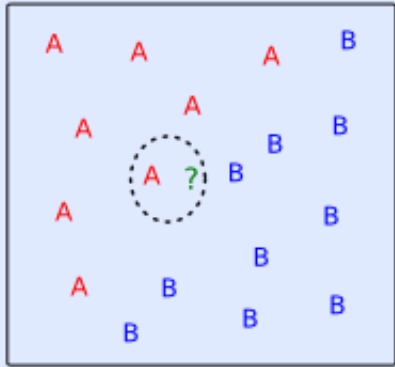
- Use labels of neighboring samples to determine label for new point



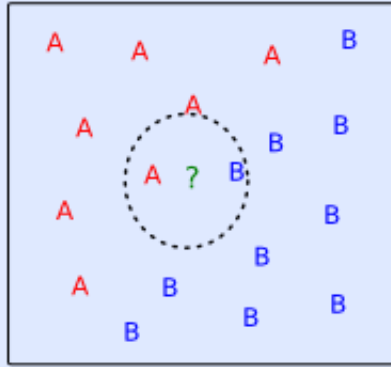
What is k?

- Value of k determines number of closest neighbors to consider

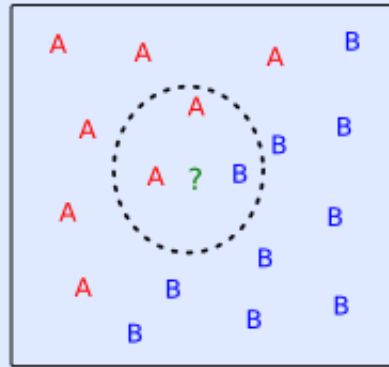
1st, 2nd, and 3rd Nearest Neighbors
of a Test Instance



1-nearest neighbor



2-nearest neighbor



3-nearest neighbor

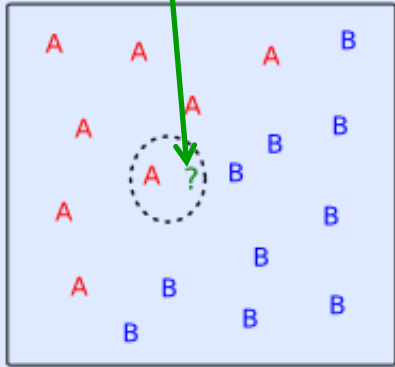
Using k Nearest Neighbors

Label='A'
(from
neighbor)

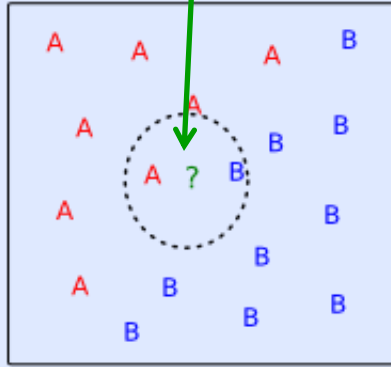
Label='B'
(random
tiebreaker)

Label='A'
(majority
vote)

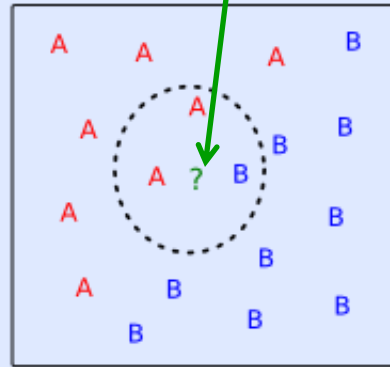
1st, 2nd, and 3rd Nearest Neighbors
of a Test Instance



1-nearest neighbor



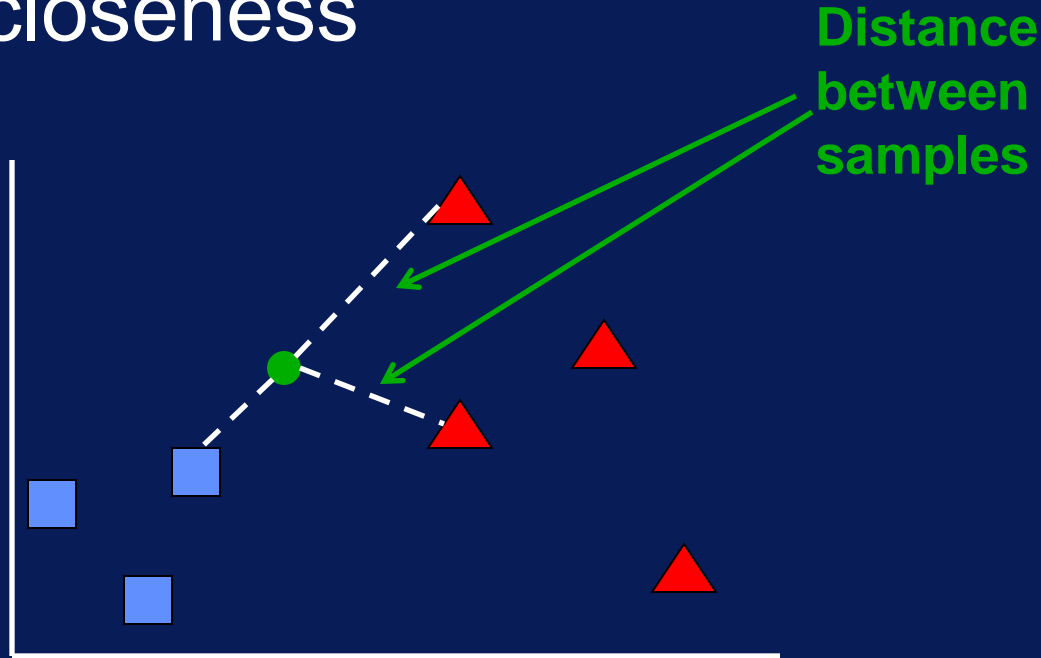
2-nearest neighbor



3-nearest neighbor

Distance Measure

- Need measure to determine “closeness”



kNN Classification

- No separate training phase
- Can generate complex decision boundaries
- Can be slow
 - Distance between new sample and all samples must be computed to classify new sample