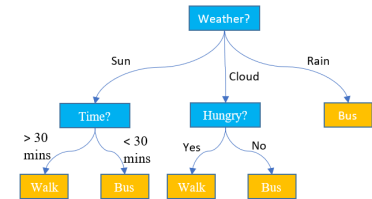
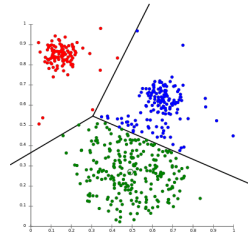
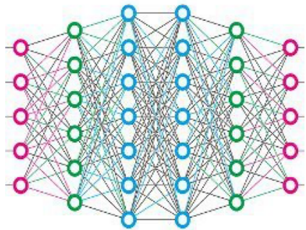
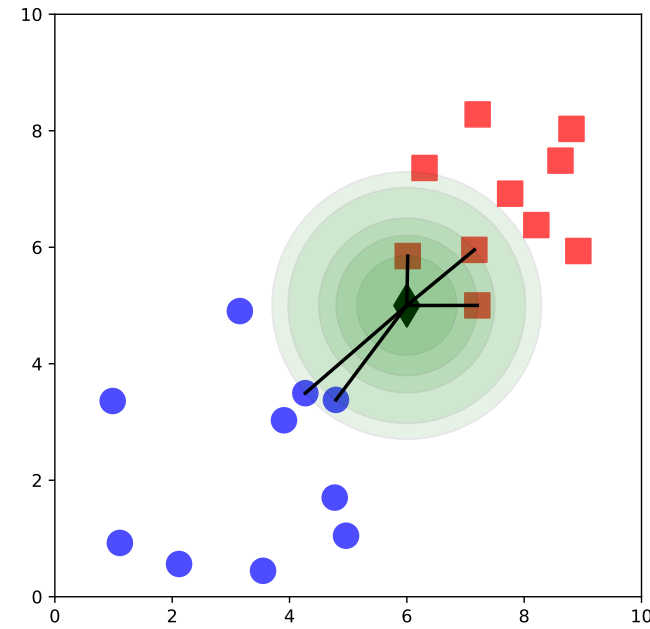


# CS 445

## Introduction to Machine Learning

### Making KNN More Efficient Comparison Classifiers

Instructor: Dr. Kevin Molloy



# Last Time

KNN uses distances to identify *similar* data in the training data and assigns the majority class from the *k* closest points/neighbors.

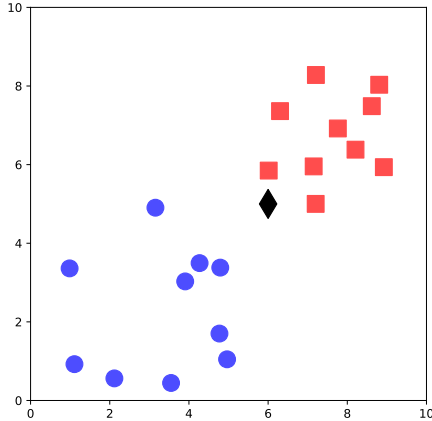
Challenges with KNN?

1. Distance computations utilize all the provided features.
  - As the number of features increases, it becomes more difficult to identify *similar* observations. Forces the issue of **feature selection**.
2. To make predictions, we need to retain the training data.
  - As we get more data (in general, this is what we want), making predictions will become slower and slower.

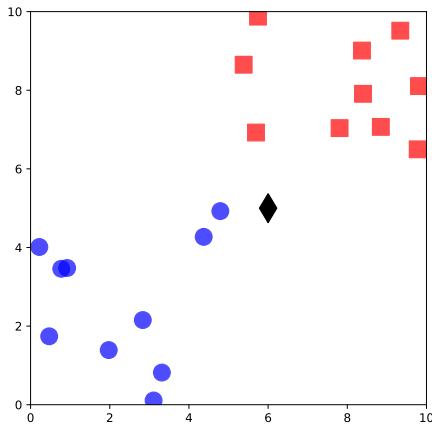
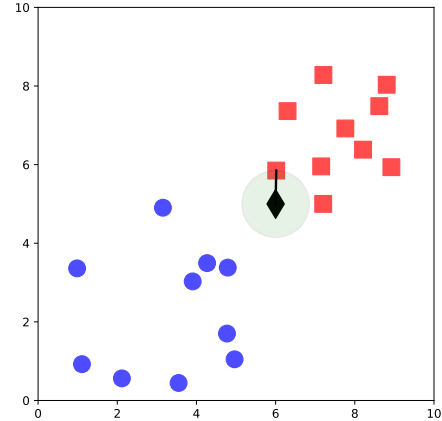
# Today's Agenda:

- Making KNN more efficient with K-D trees
- Making confidence/probability guesses with KNN
- Review that the data is the model for KNN
- Compare/constrast Decision Trees and KNN

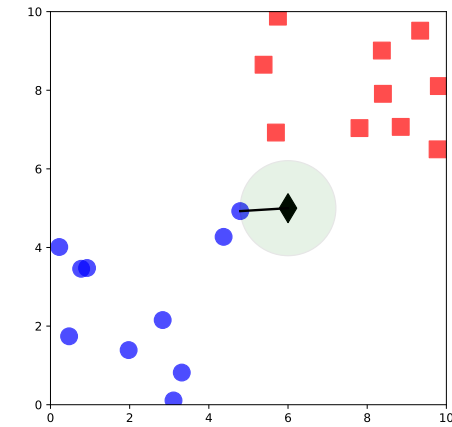
# The "Data" is the Model for KNN



Classification with Dataset A



Classification with Dataset B

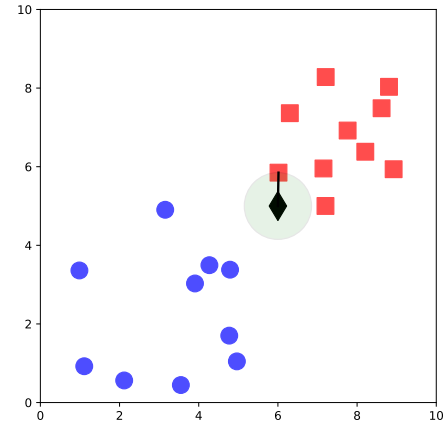


# Runtime for KNN

Building the model?      Running the model?

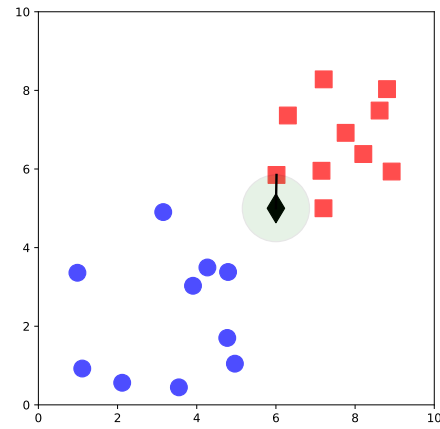
Building by default does not require anything to be done.  $O(1)$

Running predict involves identifying the  $k$  closest points. Need to measure the distance to all points  $O(n)$  or maybe  $O(nd)$  where  $d$  is the dimensions of the features used.



# KDTrees

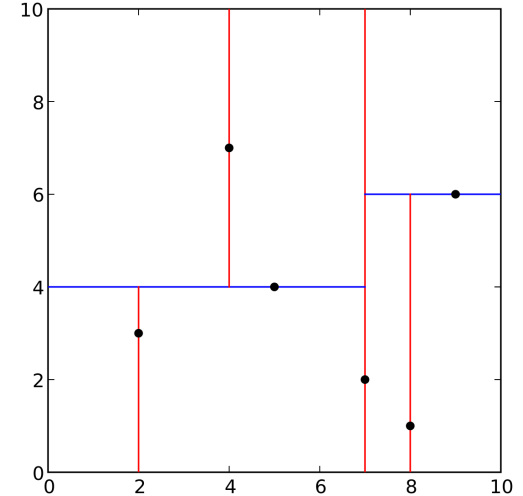
Locates nearest neighbors in some cases in  $O(\log n)$  time.



Locates nearest neighbors in some cases in  $O(\log n)$  time.

Check out this video

<https://upload.wikimedia.org/wikipedia/commons/4/48/Kdtreeogg.ogv>

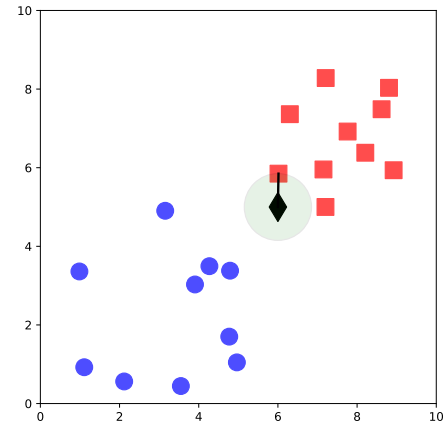


# Confidence for KNN Predictions

How can we do this?

● Vote counts

Distance voting



# KNN and Decision Trees

Compare the classifiers and report your answer to the group/class.

Group	Attribute to Compare
1	Feature Selection
2	An observation is missing a feature
3	Deals with ordinal categorical data well
4	Can deal with new data arriving continuously
5	Which classifier is more appropriate and why
6	Compare Model building and prediction performance.
All Groups	Ability to Explain the Model



# Feature Selection

Decision Trees

KNN

An observation is missing a feature

Decision Trees

KNN

Deals with ordinal categorical data well

Decision Trees

KNN

Can deal with new data arriving continuously

Decision Trees

KNN

Can deal with new data arriving continuously

Decision Trees

KNN

Can deal with new data arriving continuously

Decision Trees

KNN

# Ability to Explain the Model

Decision Trees

KNN