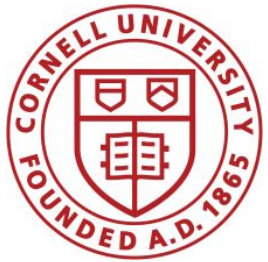



DSFA

Spring 2021

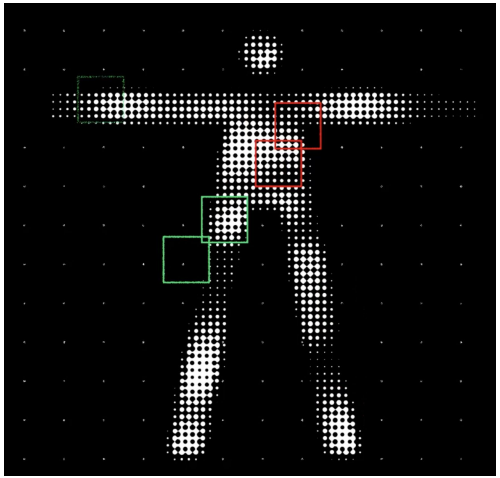
# Lecture 32

## Classification

# Announcements

- Prelim 2 regrade requests by Monday 5/3, 5PM
- HW 5 due today at 5:59PM.
- Project 3 out, Part 1 due 5/7, Part 2 due 5/14.
- Labs the week of 5/3 and 5/10 will be dedicated to Project 3.
- Final May 22, 1:30PM

# Classification

# Classification

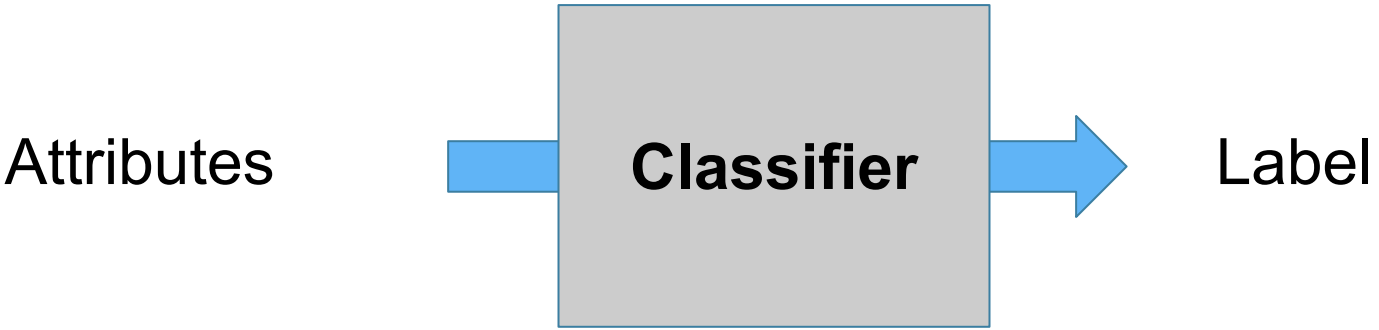
- Our study of regression:
  - One quantitative variable (x)
  - Predicts another quantitative variable (y)

- Now, classification:
  - Many quantitative variables
  - Predict a categorical variable

# Classification Terminology

- Response variable: the categorical variable we try to classify
- Classes or labels: possible values of response variable
- Binary response: 0 or 1
- Attributes: variables used to make classification

# Classifier



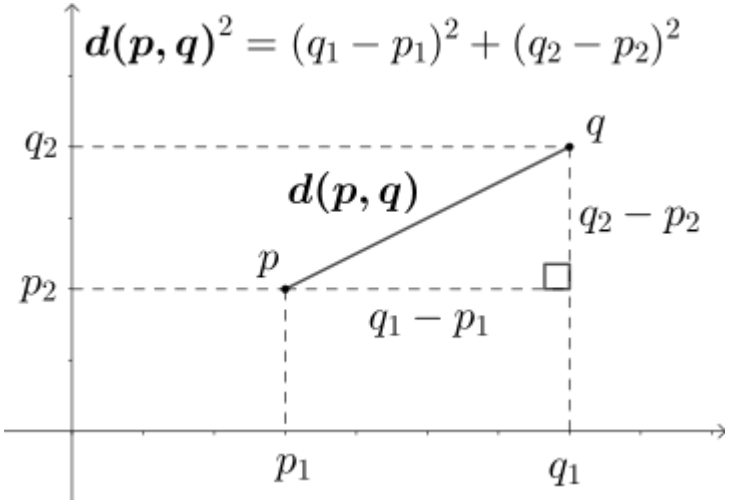

(Demo)

# Nearest Neighbor

How to classify a new individual:

- Find their nearest neighbor: the individual closest to them in the data set
- Assign the new individual the **same** label as that nearest neighbor

# Distance




(Demo)

# Nearest Neighbor recap

How to classify a new individual:

- Find their nearest neighbor: the individual closest to them in the data set
  - (We put data in standard units because scale of one attribute was so different than the other attribute--you will **not** need to do that on your proj3)
  - Compute table of distances from that individual to all other individuals
  - Sort by distance, so that closest is in the first row
- Assign the new individual the **same** label as that nearest neighbor