DSFA
Spring 2018

Lecture 34

Classification

Announcements

Correction to Lecture 33

# Classification

© Getty Images/Stock Illustration Source
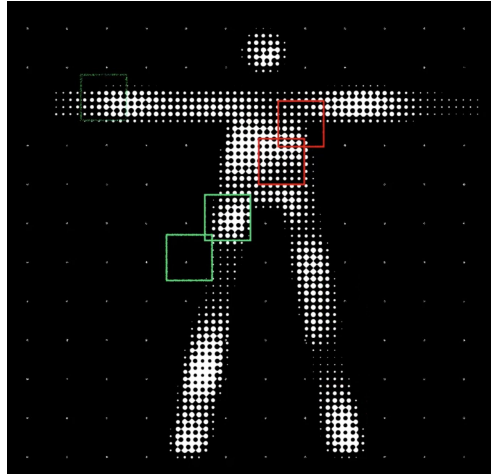
Classification

- 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
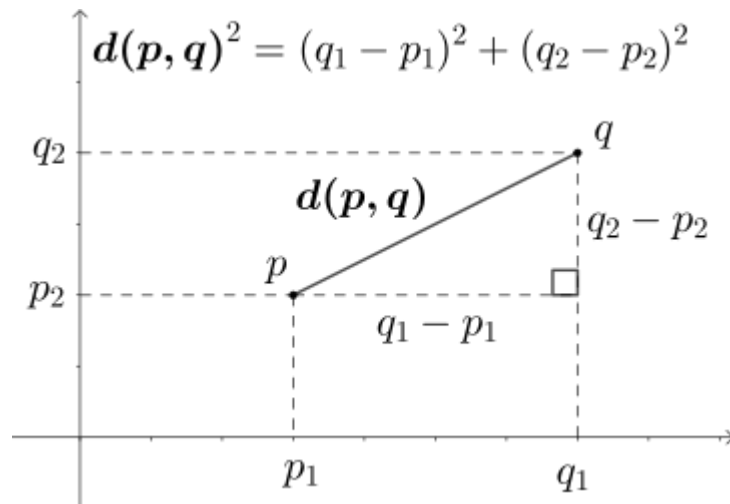- **Attributes:** variables used to make classification

Classifier

Attributes → **Classifier** → Label
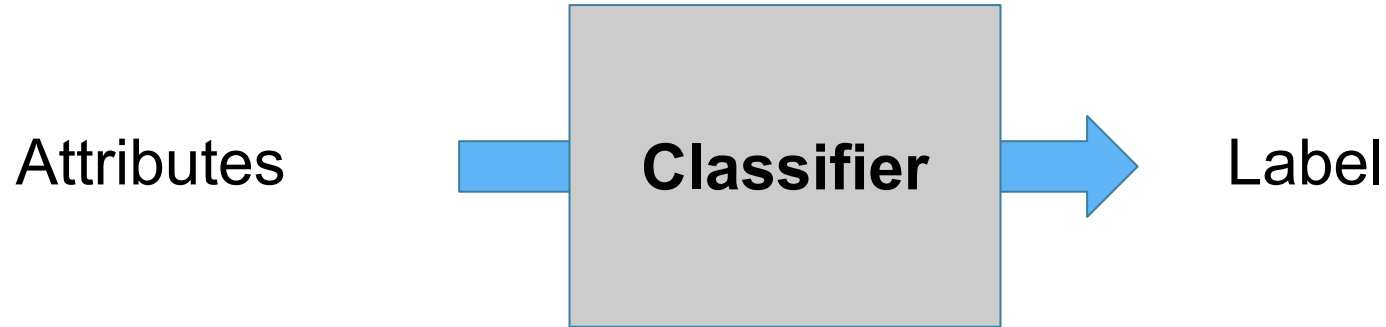
(Demo)

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

$$d(p, q)^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$

(Demo)