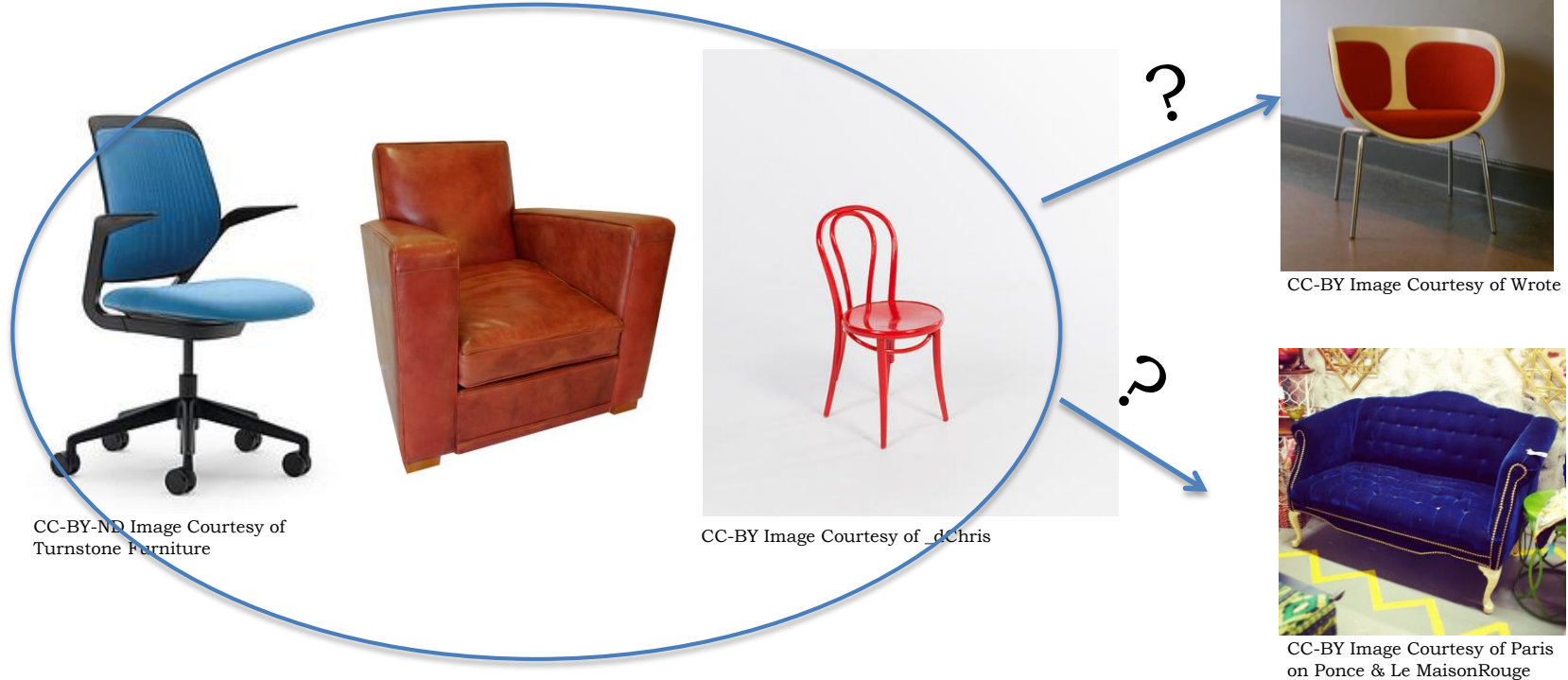


# A Quick Introduction to Machine Learning

Lecturer: John Guttag

# Machine Learning Is About Generalization



# Major Components

Method for representing the data

E.g., represent a place by its GPS coordinate

Metric for assessing goodness of the model (program)

E.g., given the GPS of a new place, how good is the model at guessing the continent

Optimization method for learning the model

E.g., linear regression

# Two Broad Classes

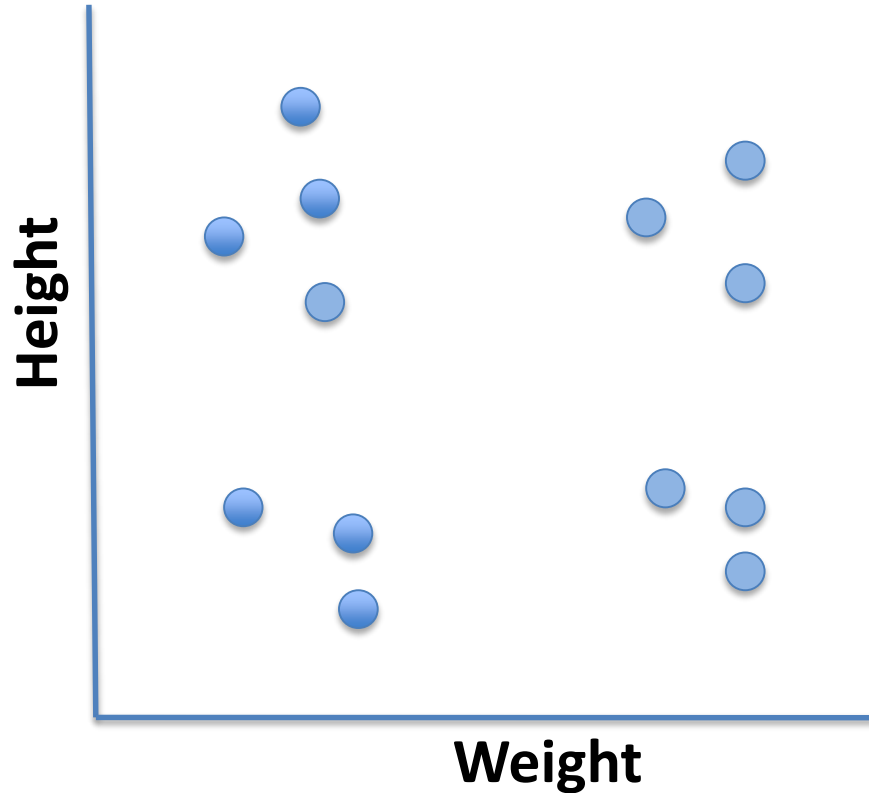
## Supervised

Given a set of feature/label pairs, find a rule that predicts the label associated with a previously unseen input

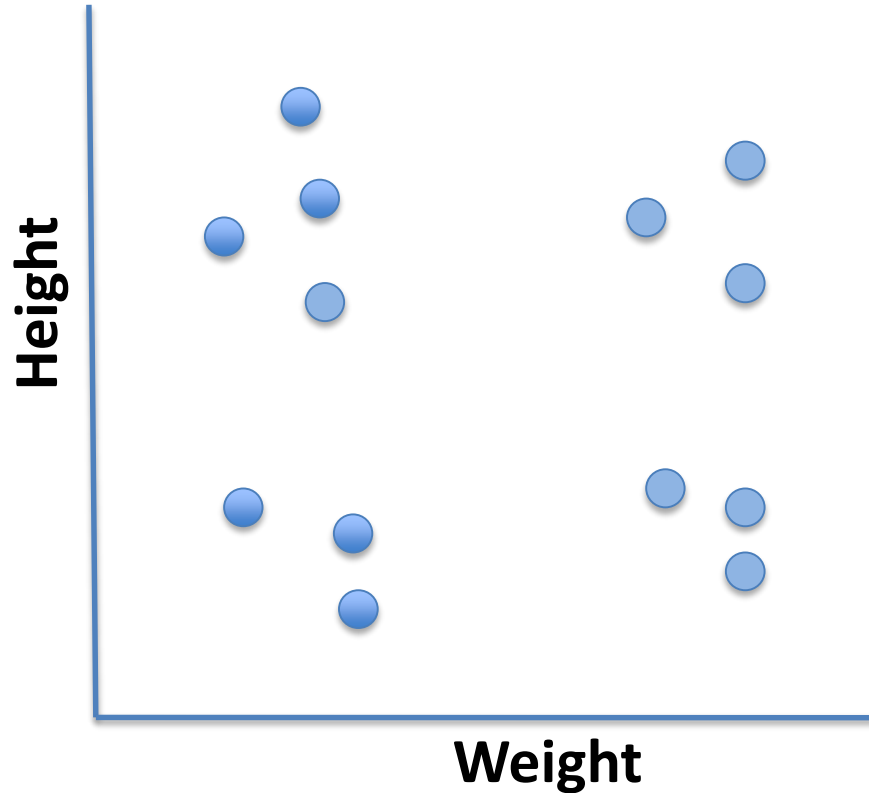
## Unsupervised

Given a set of feature vectors (without labels), group them into “natural clusters”

# Unsupervised Learning Example



# Unsupervised Learning Example



# Supervised Learning Example

Consider building a model to predict whether a person speaks English

Suppose we chose to represent a person using

- Three features

  - eye color, gender, citizenship

- Label indicating whether they speak English

# What Rule Might We Learn?


## Given

P1: <blue, male, U.S.> : true


P2: <brown, male, U.S.>: true

P3: <blue, female, Chile>: false

We probably want to learn a rule something like

For all  $x, y$ : < $x, y$ , U.S.>  speaks English

Equally likely to learn

For all  $x, z$ : < $x$ , male,  $z$ >  speaks English



# Moral

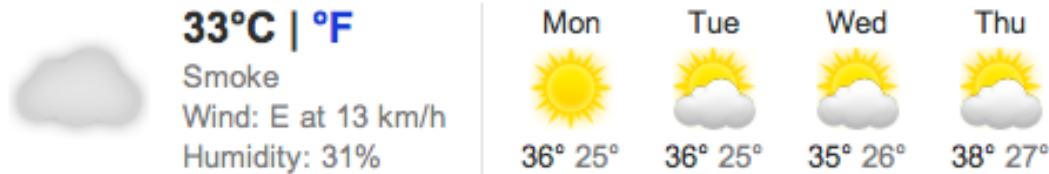
If the number of examples is small relative to the number of features, it's easy to make false generalizations

John Guttag is male and speaks English  
Eric Grimson is male and speaks English  
Therefore all males speak English

# Moral

If the number of examples is small relative to the number of features, it's easy to make false generalizations

## Weather for Mumbai, Maharashtra, India



# Moral

If the number of examples is small relative to the number of features, it's easy to make false generalizations



# Suppose We Had 100 Million Samples Drawn from Around the Globe

Probably would not learn the rule

For all  $x, z$ :  $\langle x, \text{male}, z \rangle \Rightarrow$  speaks English

But would we learn the rule

For all  $x, y$ :  $\langle x, y, \text{U.S.} \rangle \Rightarrow$  speaks English?

# It Depends

Not if we want prediction to be right 100% of time

Need to use a method that allows for training error

Want to learn something that is probably true

Remember: “All models are wrong, but some are useful”

-- George Box