INTRODUCTION TO MACHINE LEARNING

Machine Learning: What's The Challenge?

Goals of the course

- **Identify** a machine learning problem
- **Use** basic machine learning techniques
- **Think** about your data/results

What is Machine Learning?

- Construct/use algorithms that **learn** from data
- More information → Higher **performance**
- Previous solutions → Experience

Example

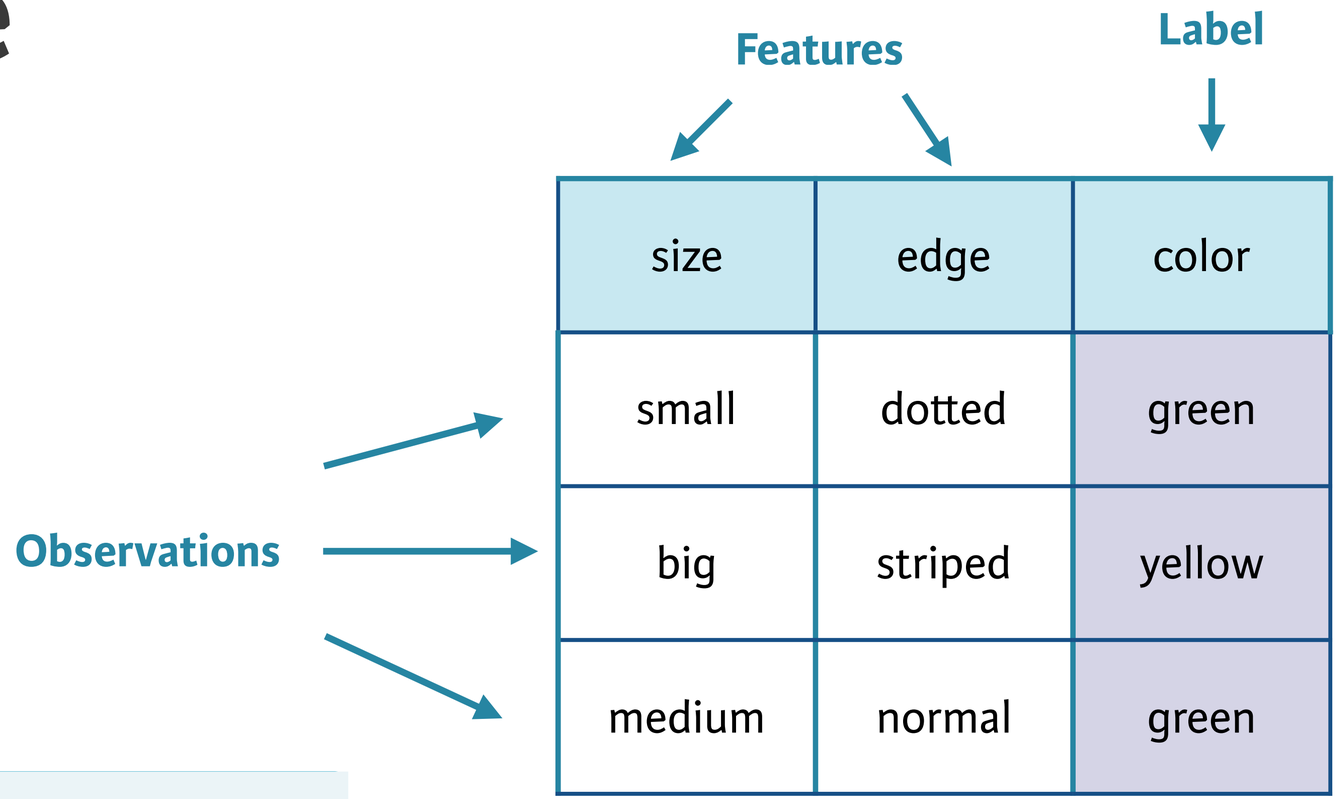
- Label squares: size and edge → **color**
- Earlier observations (labeled by humans):

- Task for computer = **label** unseen square:

- Result: right or wrong!

Input Knowledge

In example: pre-labeled squares



In **R** - use `data.frame()`

```
> squares <- data.frame(
    size = c("small", "big", "medium"),
    edge = c("dotted", "striped", "normal"),
    color = c("green", "yellow", "green"))
```

Data Frame Functions

```
> dim(squares)
```
← #Observations, #Features

```
> str(squares)
```
← Structured Overview

```
> summary(squares)
```
← Distribution Measures

Formulation

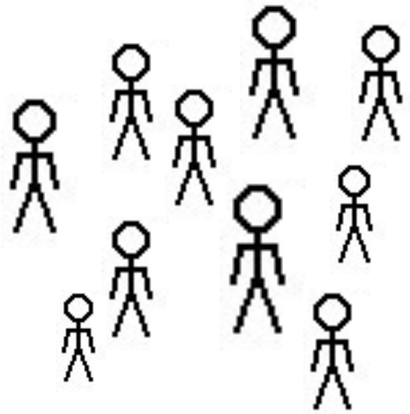
ML: What It Is Not

- Determining most occurring color
- Calculating average size

} **NOT Machine Learning**

Goal: Building models for prediction!

Regression



INPUT: Weight
OUTPUT: Height

Regression

Estimated function: $\hat{f}$

Weight

$\hat{f}$

Height

More Applications!

- Shopping basket **analysis**
- Movie **recommendation** systems
- **Decision making** for self-driving cars
- and many more!

INTRODUCTION TO MACHINE LEARNING

Let’s practice!

INTRODUCTION TO MACHINE LEARNING

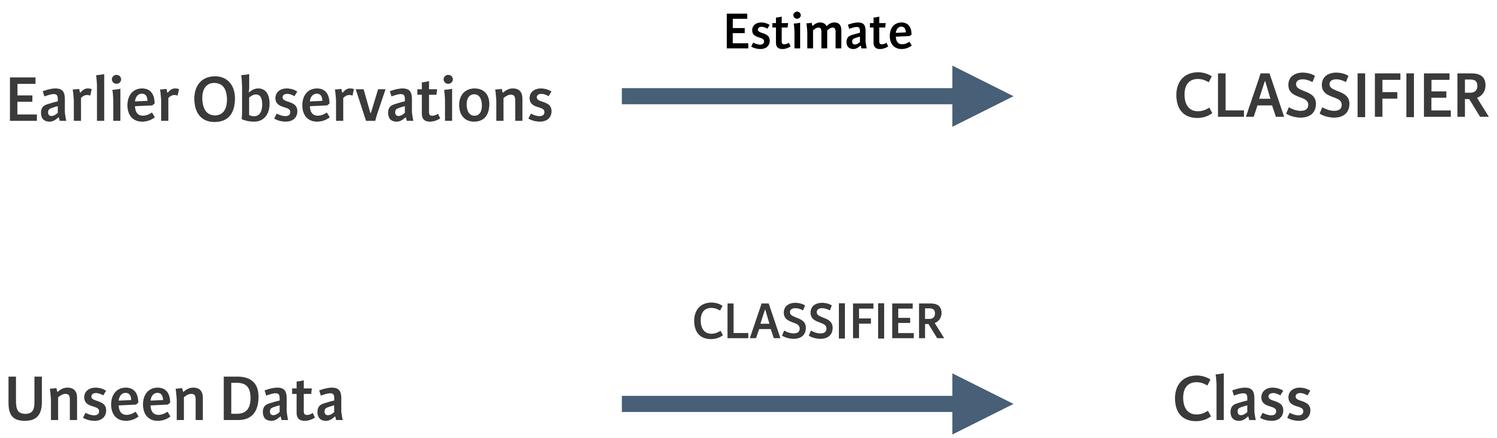
Classification
Regression
Clustering

Common ML Problems

- Classification
- Regression
- Clustering

Classification Problem

Goal: predict category of new observation



Classification Applications

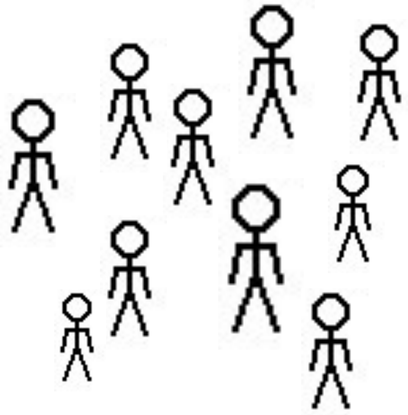
- **Medical Diagnosis** Sick and Not Sick
- **Animal Recognition** Dog, Cat and Horse

Important:

- Qualitative Output
- Predefined Classes

Regression

PREDICTORS → REGRESSION FUNCTION → RESPONSE



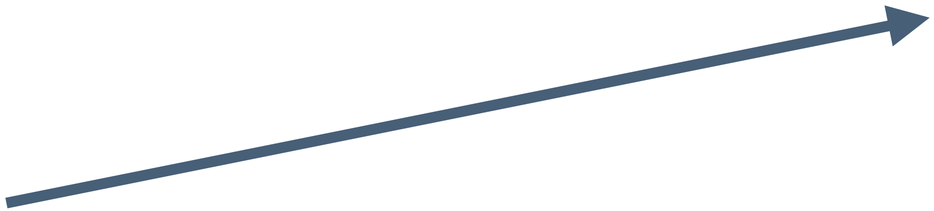
- Relationship: **Height - Weight**?
- Linear?
- Predict: Weight → **Height**

Regression Model

Fitting a **linear** function

$$\text{Height} \approx \beta_0 + \beta_1 \times \text{Weight}$$

- Predictor: Weight
- Response: Height
- Coefficients: β_0, β_1



Estimate on previous input-output

```
> lm(response ~ predictor)
```

Regression Applications

- Payments → Credit Scores
- Time → Subscriptions
- Grades → Landing a Job

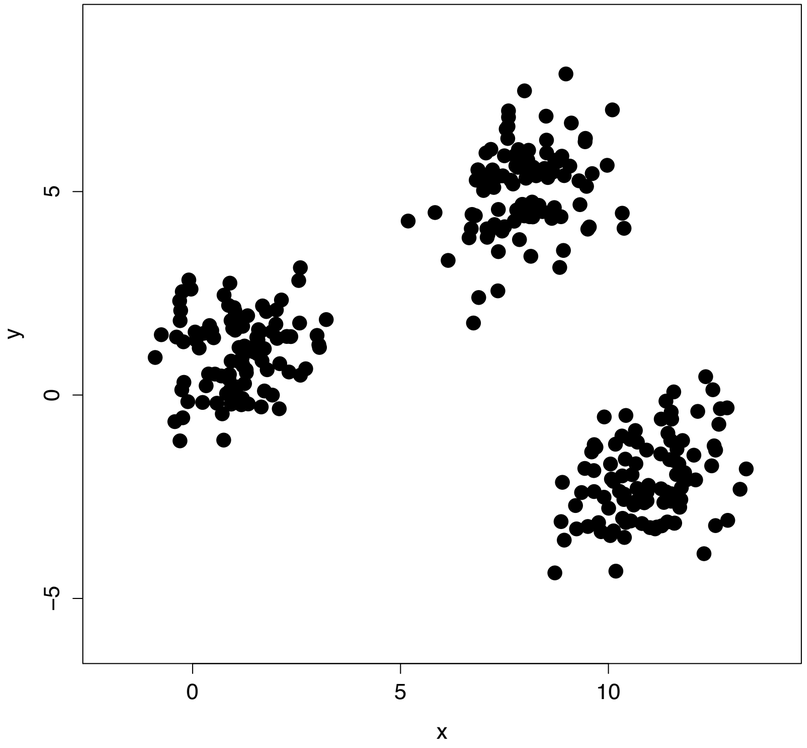
- Quantitative Output
- Previous input-output observations

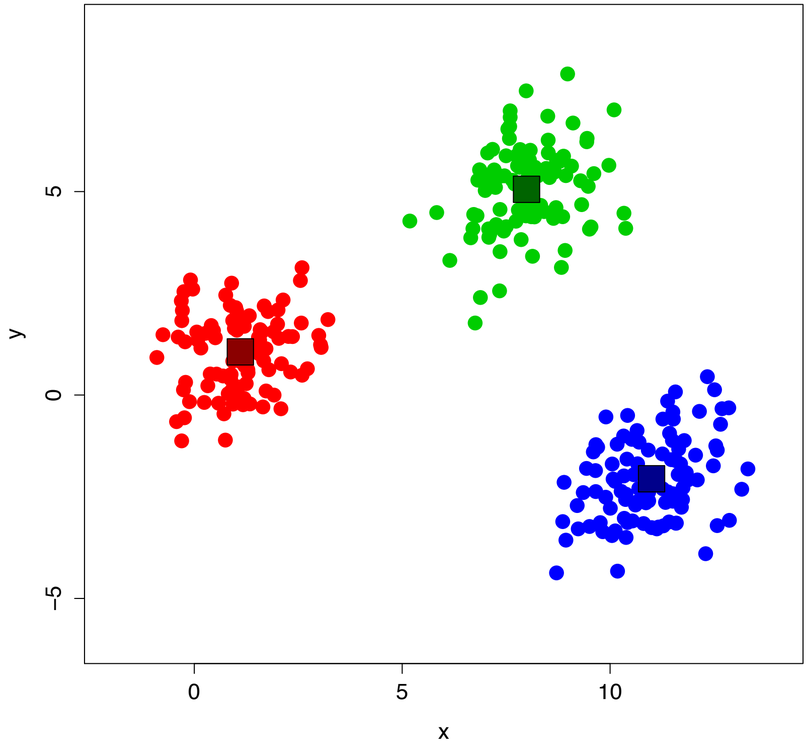
Clustering

- **Clustering**: grouping objects in clusters
  - *Similar* within cluster
  - *Dissimilar* between clusters
- **Example**: Grouping similar animal photos
  - No labels
  - No **right** or **wrong**
  - Plenty possible clusterings

k-Means

Cluster data in **k** clusters!





INTRODUCTION TO MACHINE LEARNING

Let's Practice

INTRODUCTION TO MACHINE LEARNING

Supervised vs. Unsupervised

Machine Learning Tasks

- Classification
- Regression
- Clustering

Supervised Learning

Find: function $\hat{f}$ which can be used to assign a **class** or **value** to **unseen observations.**

Given: a set of **labeled** observations

Supervised Learning

Unsupervised Learning

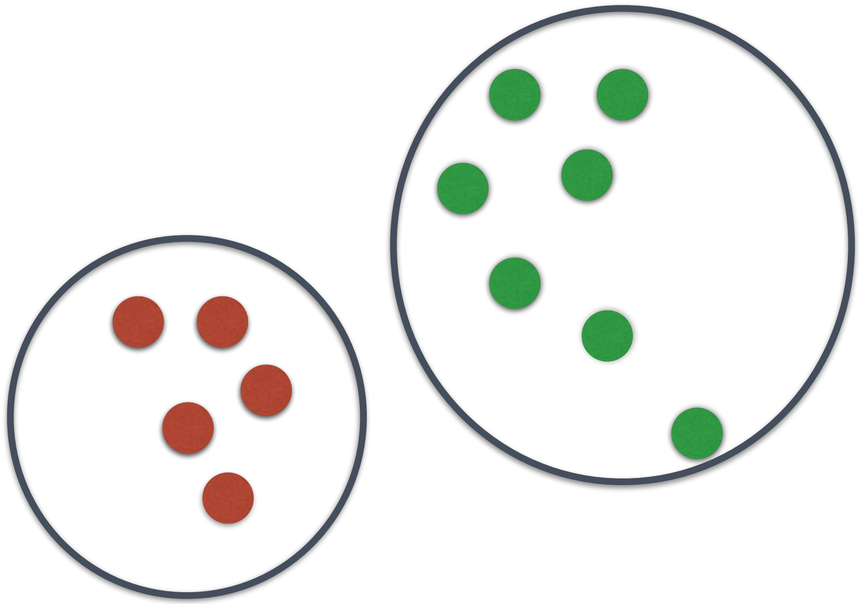
- **Labeling** can be tedious, often done by humans
- Some **techniques** don't require **labeled** data
- **Unsupervised Learning**
  - **Clustering**: find groups observation that are similar
  - Does **not** require **labeled observations**

Performance of the model

- **Supervised Learning**
  - **Compare** real labels with **predicted** labels
  - **Predictions** should be similar to **real** labels
- **Unsupervised Learning**
  - **No real labels** to compare
  - Techniques will be explained in this course

Semi-Supervised Learning

- A lot of **unlabeled observations**
- A few **labeled**
- Group similar observations using **clustering**
- Use **clustering** information and **classes** of **labeled observations** to **assign a class** to unlabelled observations
- More **labeled observations** for **supervised learning**



INTRODUCTION TO MACHINE LEARNING

Let’s practice!