

# What is Machine Learning (ML)

- Prediction
  - What will the weather be like in the next hour: sun, clouds, rain
  - Will a stock outperform the market
  - How much will a stock outperform ?

## **Informed Prediction** conditioned on information versus guessing

- What will the weather be like in the next hour ?
  - Guess: Three cases, predict each case with equal probability
  - Informed. Condition on
    - Weather now
    - Humidity
    - Season

- Will a stock outperform/underperform ?
  - Guessing: Two cases, guess each with 50% probability
  - Informed. Given the following attributes
    - Recent past performance
    - Earnings
    - Industry

## Types of prediction

- Classification: predictions come from a discrete set of values
  - Will a stock outperform or underperform the market
- Regression: predictions come from a continuous set of values
  - *How much* Will a stock outperform

Emphasis on *generalization*

- Good prediction on data never seen before (out of sample)
- Versus *explaining* in-sample data
  - Not memorization

# How this course is different

## Process versus algorithms

Machine Learning is a **process**, not a collection of algorithms !

- A methodical process to create the best prediction
- We will teach the "Recipe" for Machine Learning
  - Scientific method rather than applying an API

## Our viewpoint

- ML is an *experimental* science
  - scientific method for problem solving
  - combine engineering *and* math
- We will jump-start your experimentation: Engineering first, then math
  - Early lectures a "sprint" to get you programming and experimenting
  - Will revisit we greater mathematical basis

# Emphasis on non-traditional (for Finance) data

Non-numeric data too !

- Finance traditionally based on numeric and structured data
- Alternative data
  - Image
  - Text



We believe the future of Finance will evolve to much heavier use of unstructured data, especially non-numeric

- Images: Video, satellite
  - Forecast earnings (and hence performance) based on
    - Number of cars in parking lot
    - Activity of supply chain (number of shipments or components)

- Text/speech
  - Was the sentiment of a news article about the company Positive or Negative
  - Did the tone of the CFO's conference call reveal lack of conviction ?

# ML and Finance, rather than ML for Finance

Machine Learning is a skill applied to many domains, not just Finance

- Transferable skill set
- Innovation originates outside and migrates to Finance
- Finance: historically numeric data
  - Wide opportunity set (for Finance) in non-numeric data (Images, Text) which we will study
- We will learn ML with Finance as examples