# Course Overview

Professor Ameet Talwalkar

# Outline

# What is machine learning?

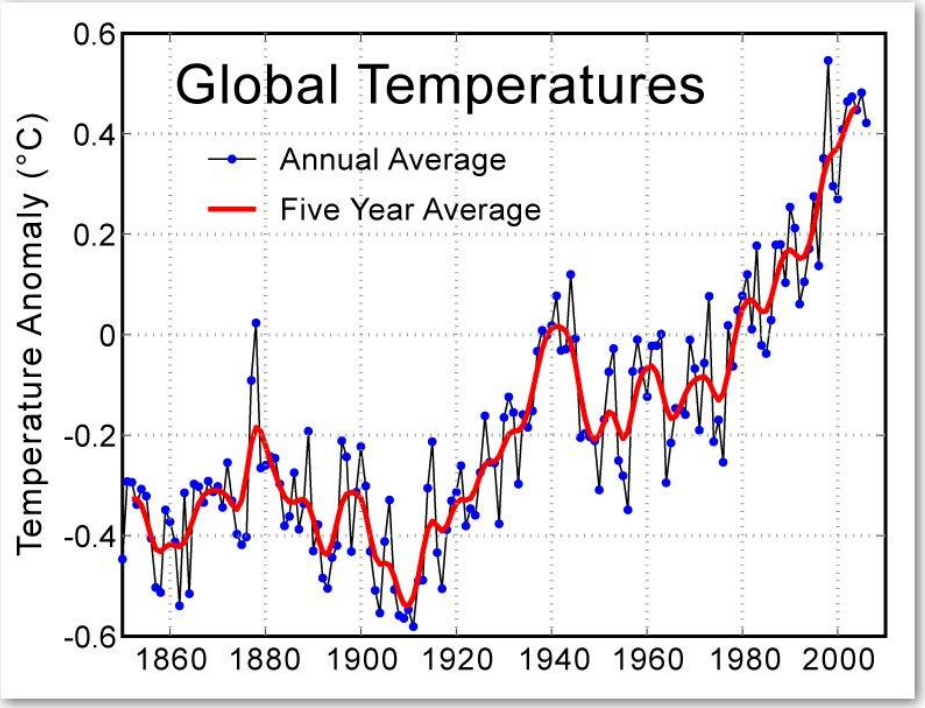
**One possible definition**[1]

a set of methods that can automatically *detect patterns* in data, and then use the uncovered patterns to *predict future data*, or to perform other kinds of decision making *under uncertainty*

cf. Murphy's book

# Example: detect patterns

**How the temperature has been changing in the last 140 years?**



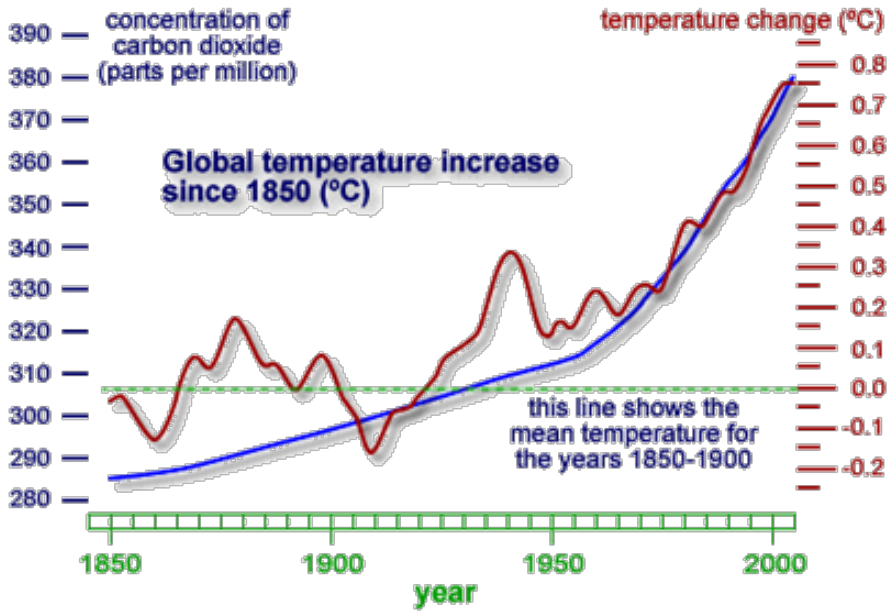

## Patterns

- We see repeated periods of fluctuation
- General trend is that temperatures are rising

# How do we describe the pattern?

# How do we describe the pattern?

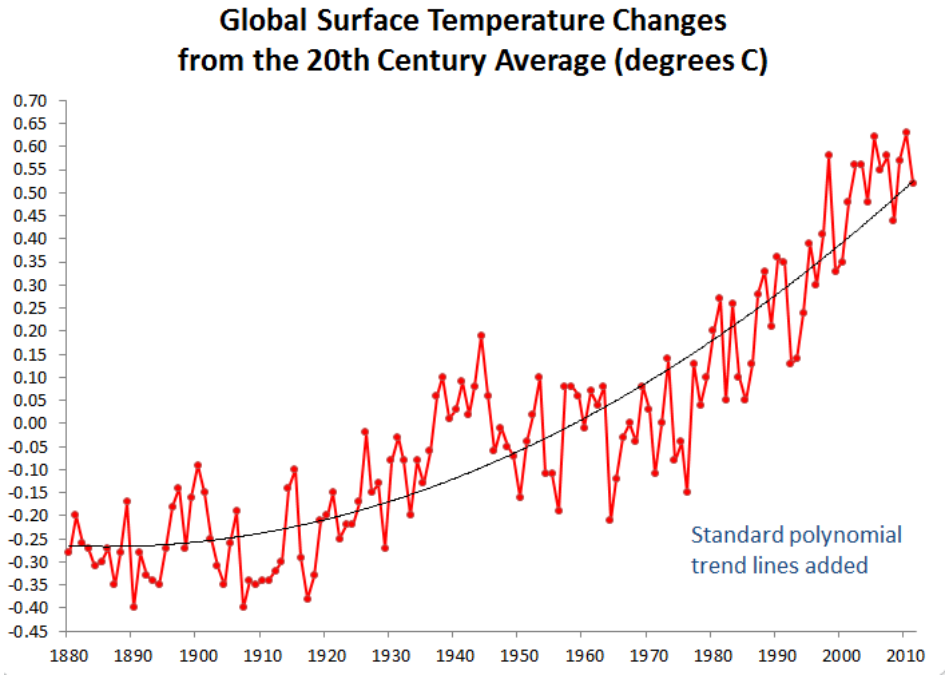
**Build a model: fit the data with a polynomial function**




- The model is not accurate for individual years
- But overall, the model captures the major trend

# Predicting future

## What is temperature of 2017?




- This particular polynomial model is not exactly accurate for that specific year, but it is pretty close

# What we have learned from this example?

**Key ingredients in the machine learning task**

# What we have learned from this example?

**Key ingredients in the machine learning task**

- Data: collected from past observations (*training data*)

# What we have learned from this example?

**Key ingredients in the machine learning task**

- Data: collected from past observations (*training data*)
- Modeling: devised to capture the patterns in the data
  - The model does not have to be true — as long as it is close, it is useful
  - We should tolerate randomness and mistakes — many interesting things are stochastic by nature.

# What we have learned from this example?

**Key ingredients in the machine learning task**

- Data: collected from past observations (*training data*)
- Modeling: devised to capture the patterns in the data
  - The model does not have to be true — as long as it is close, it is useful
  - We should tolerate randomness and mistakes — many interesting things are stochastic by nature.
- Prediction: apply the model to forecast what is going to happen in future

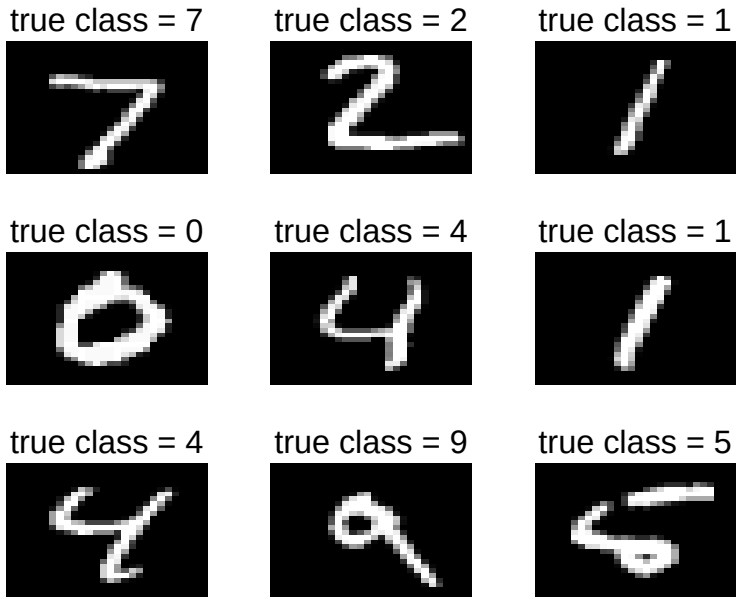
# A rich history of applying statistical learning methods

**Recognizing flowers (by R. Fisher, 1936)**

Types of Iris: setosa, versicolor, and virginica

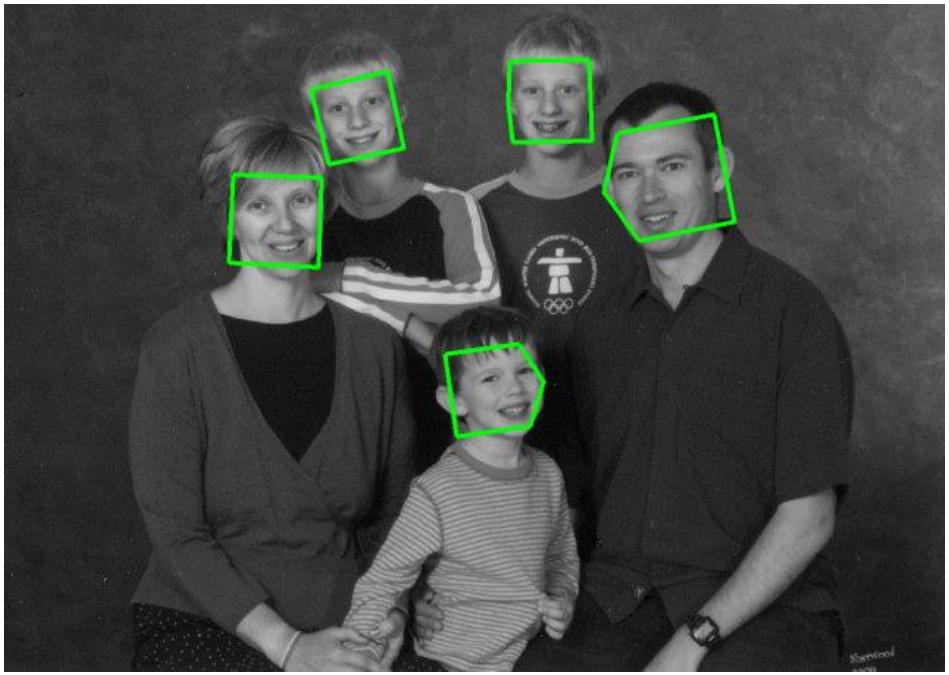
# Huge success 20 years ago

**Recognizing handwritten zipcodes and checks (AT&T Labs, circa late 1990s)**



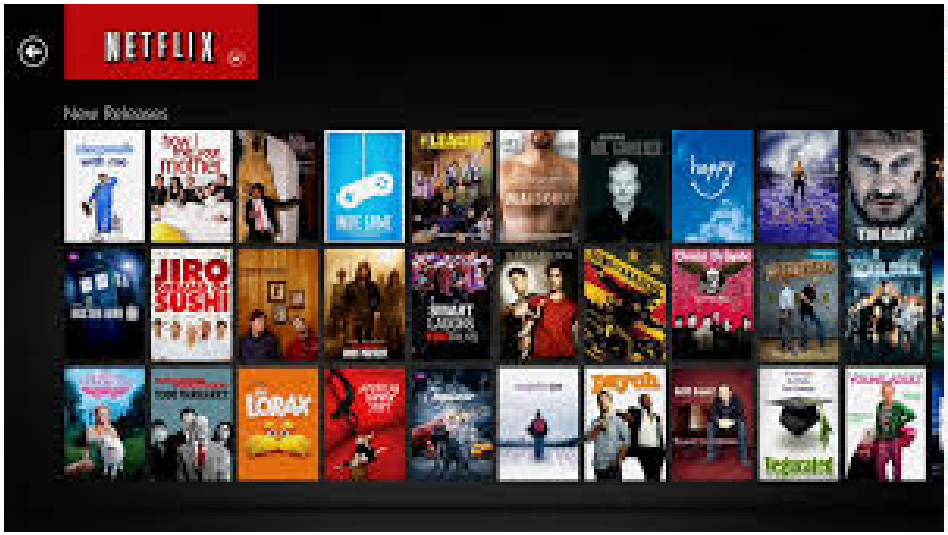
# More modern ones, in your social life

## Recognizing your friends on Facebook



# Learn your preferences

## Recommending what you might like




# Why is machine learning so popular?

# Why is machine learning so popular?

- Data
  - Flood of data from various sensors leads to several high-impact applications
  - e.g., cell phones, internet applications, scientific studies

- Computing
  - Powerful and cheaply available computing resources enables efficient storage / processing / analysis of this data
  - e.g., cloud computing, GPUs, cell phones

# Some Applications

- Consumer applications:
  - speech recognition, information retrieval and search, email and document classification, stock price prediction, object recognition, product recommendation, $\cdots$
  - Highly desirable expertise from industry: Google, Facebook, Microsoft, Yahoo, Twitter, IBM, LinkedIn, Amazon, $\cdots$
- Scientific applications:
  - Biology and genetics: identify disease-causing genes and gene networks
  - Climate science: predicting global warming trends
  - Social science: social network analysis; social media analysis
  - Business and finance: marketing, operation research
  - Emerging ones: healthcare, energy, $\cdots$

# What is in machine learning?

## Different flavors of learning problems

- Supervised learning: make prediction given labeled training observations, e.g., Spam detection, Iris

# What is in machine learning?

## Different flavors of learning problems

- Supervised learning: make prediction given labeled training observations, e.g., Spam detection, Iris
- Unsupervised learning: Discover hidden and latent patterns in data; data exploration, e.g., topic modelling in text data

# What is in machine learning?

## Different flavors of learning problems

- Supervised learning: make prediction given labeled training observations, e.g., Spam detection, Iris
- Unsupervised learning: Discover hidden and latent patterns in data; data exploration, e.g., topic modelling in text data
- Reinforcement learning: act optimally (or at least well) under uncertainty, e.g., defining a robot's behavior with the world based on the feedback (rewards / punishments) it receives from each action

# What is in machine learning?

## Different flavors of learning problems

- Supervised learning: make prediction given labeled training observations, e.g., Spam detection, Iris
- Unsupervised learning: Discover hidden and latent patterns in data; data exploration, e.g., topic modelling in text data
- Reinforcement learning: act optimally (or at least well) under uncertainty, e.g., defining a robot's behavior with the world based on the feedback (rewards / punishments) it receives from each action
- Many other paradigms

## The focus and goal of this course

- Supervised learning (majority of the course)
- Unsupervised learning (last week of course)

# How to grasp machine learning well

## Pillars to machine learning

- Statistics
- Linear Algebra
- Optimization
- Algorithms

## Resources to study them

- Suggested Reading:
  - Chapter 2 of MLAPA book
  - Linear Algebra Review and Reference by Zico Kolter and Chuong Do (`http://www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf`)
  - Convex Optimation Review by Zico Kolter and Honglak Lee (`http://www.cs.cmu.edu/~./15381/slides/cvxopt.pdf`)
- Wikipedia (some information might not be 100% accurate, though)

# Outline

# Quick Polls

- Undergrad / Masters / PhD?

# Quick Polls

- Undergrad / Masters / PhD?
- Computer Science / Other Engineering / Other?

# Quick Polls

- Undergrad / Masters / PhD?
- Computer Science / Other Engineering / Other?
- Registered / Waiting List / Hoping to Register / Other?

# Discussion Section

- There will only be one discussion section
  - Friday 12:00p - 1:50pm, Haines 118
  - Led by Chris Wu

- The second section has been been cancelled, but one of the TAs will be holding office hours at this time
  - Friday 11:00a - noon

# Registration / PTEs

- Course is currently full and I can't increase class size
- **I am not giving out PTEs** until after HW1 submission date

# Registration / PTEs

- Course is currently full and I can't increase class size
- **I am not giving out PTEs** until after HW1 submission date
- I expect several students will drop the course
  - ML is very popular, and many students are interested but don't have mathematical maturity for this graduate level material
  - Last year's attrition rate much higher than typical grad-level class
  - Several students dropped late in the quarter, thus preventing other students from joining

# Registration / PTEs (cont)

- I believe / hope that all qualified students will be able to enroll
- Today's math quiz will hopefully mitigate attrition later in the course
  - Representative of mathematical concepts you are excepted to know
  - This is a course requirement
  - Graded to assess your background (but not part of final grade)
  - We may contact students who perform poorly

# Registration / PTEs (cont)

- I believe / hope that all qualified students will be able to enroll
- Today's math quiz will hopefully mitigate attrition later in the course
  - Representative of mathematical concepts you are excepted to know
  - This is a course requirement
  - Graded to assess your background (but not part of final grade)
  - We may contact students who perform poorly
- Be honest / realistic with yourself about your background
  - It's better for you, me, and your classmates to drop the course now rather than a month from now
- If you're not registered, I'd encourage you to stay patient
  - Priority will be given to students based on their scores on the math quiz, first problem set, and CS affiliation

# Course Logistics

Let's go to the course website:

`http://cs.ucla.edu/~ameet/teaching/winter17/cs260/index.html`

# Any questions?