

Day 1: Introduction to Machine Learning

Summer STEM: Machine Learning

Nikola Janjušević
Akshaj Kumar Veldanda
Jacky Yuan
Tejaishwarya Gagadam

NYU Tandon School of Engineering

June 13, 2019

Learning Objectives

- What is Machine Learning?
- What is Regression?
- What is Classification?
- Why is Machine Learning gaining so much importance?
- Where do I find data, how is it represented?
- How do we use Python to represent, manipulate, and visualize data?
- How do we use vectors and matrices to manipulate data?
- How do computers understand continuous data? How do we visualize it?
- How do we use random numbers in programming and modelling noise?

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Teacher Introductions

Nikola

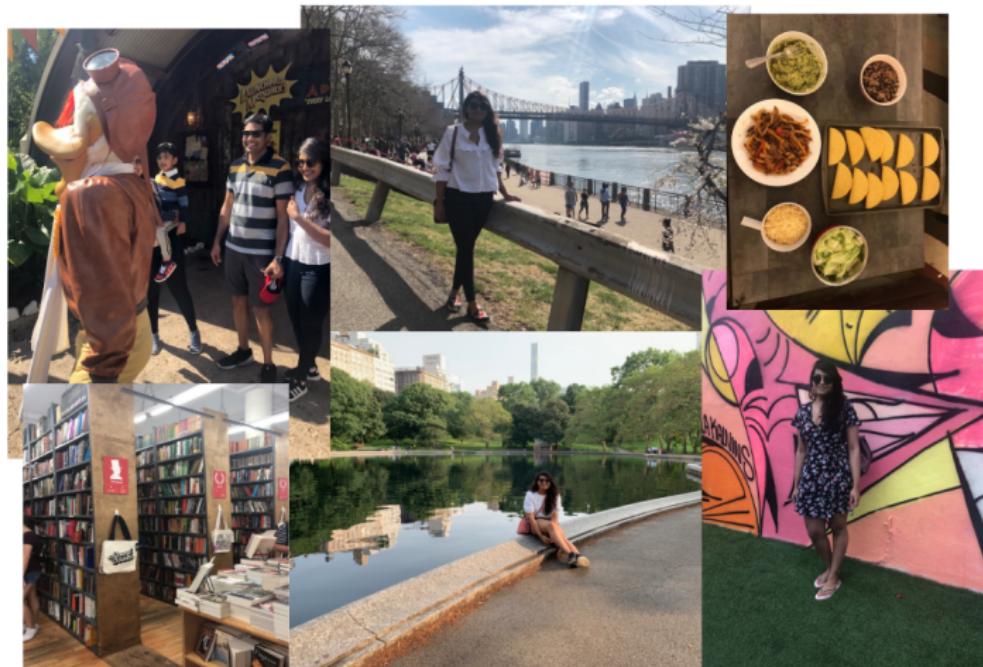


Akshaj



Teacher Introductions

Aishwarya

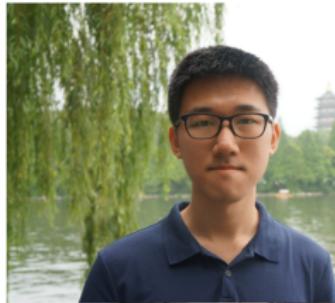


Nikola, Akshaj, Jacky, Aishwarya

Day 1: Introduction to Machine Learning

Teacher Introductions

Jacky



Tell the class about yourself

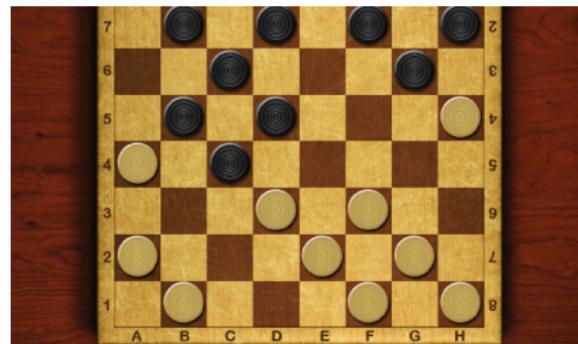
- Write down the following pieces of information on your notecard:
 - Name
 - Grade
 - Where are you from?
 - What do you want to get out of this class?
 - What time did you wake up this morning?
 - Where did you travel from?
 - What mode of transport did you take?
 - How long did it take you? use Google Maps if unsure
 - One rule you'd like to propose for the classroom
- Stand in front of the class and present what you've written down
- We'll visualize this data in python later today
 - Link to excel sheet here

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Definition

- Arthur Samuel (1959): Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.



Another Example: Digit Recognition



- Challenges with expert approach
 - Simple expert rule breaks down in practice
 - Difficult to translate our knowledge into code
- Machine Learning approach
 - Learned systems do very well on image recognition problems
 - This is what you will learn in this course

```
def classify(image):  
    ...  
    nv = count_vert_lines(image)  
    nh = count_horiz_lines(image)  
    ...  
  
    if (nv == 1) and (nh == 1):  
        digit = 7  
    ...  
  
    return digit
```

Modern Definition

Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Example: playing checkers.

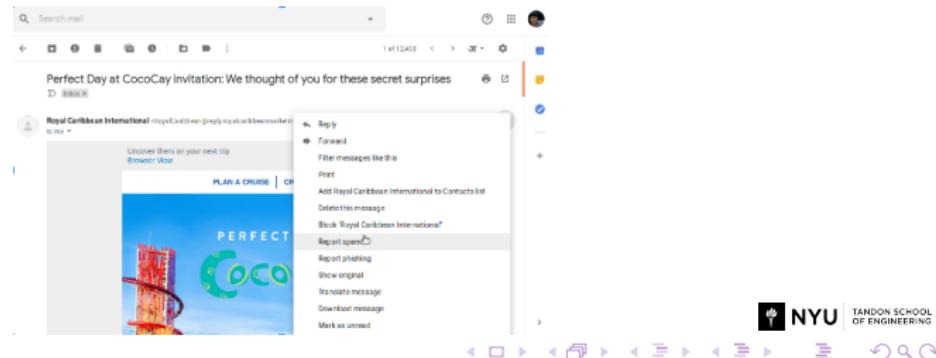
- E = the experience of playing many games of checkers
- T = the task of playing checkers.
- P = the probability that the program will win the next game.

Question

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam.

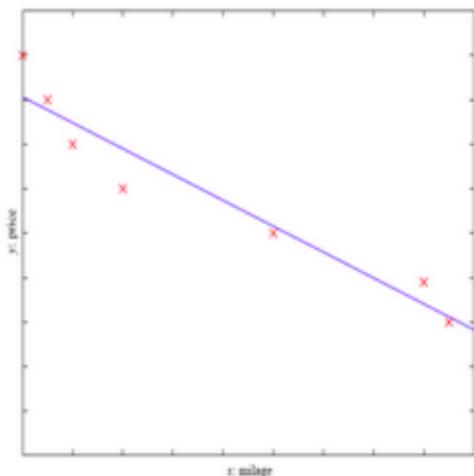
What is the task, experience and performance in this setting?

- a) Classifying emails as spam or not spam
- b) Watching you label emails spam or not spam
- c) The number of emails correctly classified as spam or not spam



What is Regression?

- Target variable is continuous-valued
- Example
 - Predict $y = \text{price}$ of a car
 - From $x = \text{mileage, size, horsepower}$
 - Can use multiple predictors
- Assume some form of mapping
 - Ex: Linear mapping: $y = \beta_0 + \beta_1 x$
 - Find parameter β_0, β_1 from data



What is Classification?

- Object classification from images
- Determine what objects are in an image
- Features
 - Pixels of the image
- Target
 - Classification label
- Learn a model / function from features to target
 - Use examples of labeled images



Outline

- 1 Teacher and Student Introductions
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?
- 3 Course Outline
- 4 Dealing with Data
- 5 Manipulating Data in Math and Code
- 6 Some Statistics
- 7 Lab: Generating a Sine-Wave
- 8 Some Probability
- 9 Lab: Generating a Noisy Sine-Wave
- 10 Artificial Intelligence and Machine Learning
- 11 Why the Hype?

Course Outline

- Day 1: Intro to ML
- Day 2: Linear Regression
- Day 3: Generalization Error
- Day 4: Linear Classification
- Day 5: Mini-Project Competition & Presentations
- Day 6: Neural Networks
- Day 7: Deep Learning & Convolutional Neural Networks
- Day 8: Applications of CNNs
- Day 9: Final Projects
- Day 10: Final Projects & Presentations

Course Format, Website, Resources

- Course Website: github.com/nikopj/SummerML
 - Contains lecture slides, code notebooks, and datasets
 - Slides posted after lecture, student code notebooks and datasets the morning of
- We're programming in Python via Google Colab
 - No installation required
- We'll give additional resources at the end of each day based on student interest

Github & Colab

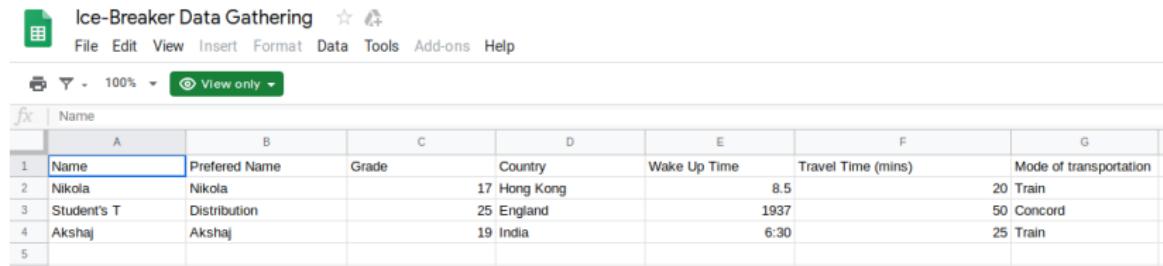
- Github is an online service for sharing code repositories and collaborating on projects
- Google Colab runs Python code on a server or your local machine
 - Works just like Jupyter Notebook
 - Write separate code blocks and markdown blocks

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Looking at our "ice-breaker" data in spreadsheets

- Columns have labels in the first row
- Collected data (numbers, words) follow below
- Let's export it to a Comma-Separated Values (CSV) file and open it
- Demo on Github.



The screenshot shows a Google Sheets document titled "Ice-Breaker Data Gathering". The menu bar includes File, Edit, View, Insert, Format, Data, Tools, Add-ons, and Help. The toolbar shows a zoom level of 100% and a "View only" button. The spreadsheet has a header row with column labels A through G. The data starts with a row for "Name" and "Preferred Name", followed by rows for "Grade", "Country", "Wake Up Time", "Travel Time (mins)", and "Mode of transportation". The data entries are: Nikola (Grade 17, Hong Kong, 8.5, Train), Student's T (Grade 25, England, 1937, Concord), and Akshaj (Grade 19, India, 6:30, Train).

	Name	Preferred Name	Grade	Country	Wake Up Time	Travel Time (mins)	Mode of transportation
1	Nikola	Nikola		17 Hong Kong	8.5		Train
2	Student's T	Distribution		25 England	1937		Concord
3	Akshaj	Akshaj		19 India	6:30		Train
4							
5							

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Vectors

- A **vector** is an n-tuple of numbers or symbols
 - Ex: integer 4-tuple: $\mathbf{v} = (1, 5, 2, 9)$
- Vectors of the same size may be added together, element-wise
 - Ex: $\mathbf{u} = (3, 6, 0, -5)$
 $\mathbf{v} + \mathbf{u} = (1 + 3, 5 + 6, 2 + 0, 9 + (-5)) = (4, 11, 2, 4)$
- Vectors may be scaled by a number, element-wise
 - Ex: $3\mathbf{v} = (3 \times 1, 3 \times 5, 3 \times 2, 3 \times 9) = (3, 15, 6, 27)$
- **Dot-Product:** sum of element-wise products of two vectors
 - Ex:
$$\mathbf{u} \cdot \mathbf{v} = 3 \times 1 + 6 \times 5 + 0 \times 2 + (-5) \times 9 = 3 + 30 - 45 = -12$$
- Length of a Vector (L₂ Norm): $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$
 - Ex: $|\mathbf{v}| = \sqrt{1 + 25 + 4 + 81} = \sqrt{111}$

Matrices

- A **matrix** is a rectangular array of numbers or symbols arranged in rows and columns
 - Ex: 2 by 3 matrix, $M = \begin{bmatrix} 1 & 9 & -12 \\ 15 & -2 & 0 \end{bmatrix}$
- Matrices of the same shape may be added together, element-wise
 - Ex: $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 8 \\ 7 & 11 \end{bmatrix}, A + B = \begin{bmatrix} 1 & 9 \\ 9 & 12 \end{bmatrix}$
- Matrices may be scaled, element-wise
 - Ex: $aB = \begin{bmatrix} 0 & 8a \\ 7a & 11a \end{bmatrix}$, where a is a scalar

More Matrix Operations

- Matrices may be multiplied together provided their shapes meet the criteria
 - Criteria: for AB to be valid, # cols of A must equal the # rows of B
 - Result is a matrix with shape (# rows A, # cols B)
 - Ex: A is (2×3) and B is (3×5) , $C = AB$ is (2×5)
 - Example on board
- **Transpose:** A^T swaps the rows and columns of matrix A
- **Inverse:** A^{-1} satisfies the equation $AA^{-1} = A^{-1}A = \mathbb{I}$
 - Square matrices only!

Vectors and Matrices

- We may consider a vector as a matrix

- **Row Vector:** shape $(1 \times N)$

Ex: $\mathbf{v} = [1 \ 5 \ 2 \ 9]$

- **Column Vector:** shape $(N \times 1)$

$$\text{Ex: } \mathbf{v} = \begin{bmatrix} 1 \\ 5 \\ 2 \\ 9 \end{bmatrix} = [1 \ 5 \ 2 \ 9]^T$$

- We'll consider vectors as column vectors by default

- Vector-Matrix Multiplication: on board
- Dot Product:

Vectors and Matrices

- We may consider a vector as a matrix

- **Row Vector:** shape $(1 \times N)$

Ex: $\mathbf{v} = [1 \ 5 \ 2 \ 9]$

- **Column Vector:** shape $(N \times 1)$

$$\text{Ex: } \mathbf{v} = \begin{bmatrix} 1 \\ 5 \\ 2 \\ 9 \end{bmatrix} = [1 \ 5 \ 2 \ 9]^T$$

- We'll consider vectors as column vectors by default

- Vector-Matrix Multiplication: on board
- Dot Product: $\mathbf{v}^T \mathbf{v}$

Numpy Demo

- Dealing with vectors and matrices using the Python Module, Numpy

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Sample Mean and Variance

- Observations of a random variable

Sample Mean and Variance

- Observations of a random variable
 - Ex: We roll a 6 sided die 10 times with outcomes
$$X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$$

Sample Mean and Variance

- Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes
 $X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$
- What is the average outcome?

Sample Mean and Variance

- Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes
 $X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$
- What is the average outcome?

- **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$

Sample Mean and Variance

- Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes

$$X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$$

- What is the average outcome?

- **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$

- How do our samples vary around the average?

Sample Mean and Variance

■ Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes

$$X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$$

- What is the average outcome?

- **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$

- How do our samples vary around the average?

- **Sample Variance:** $\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$

Sample Mean and Variance

- Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes

$$X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$$

- What is the average outcome?

- **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$

- How do our samples vary around the average?

- **Sample Variance:** $\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$

- **Standard Deviation:** σ_x

Sample Mean and Variance

■ Observations of a random variable

- Ex: We roll a 6 sided die 10 times with outcomes

$$X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$$

- What is the average outcome?

- **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$

- How do our samples vary around the average?

- **Sample Variance:** $\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$

- **Standard Deviation:** σ_x

- *Weighing the deviation of the mean by the magnitude of the deviation*

Sample Mean and Variance

- Observations of a random variable
 - Ex: We roll a 6 sided die 10 times with outcomes $X = (x_1, x_2, \dots, x_{10}) = (1, 5, 2, 3, 4, 4, 6, 2, 3, 5)$
 - What is the average outcome?
 - **Sample Mean:** $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{35}{10} = 3.5$
 - How do our samples vary around the average?
 - **Sample Variance:** $\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$
 - **Standard Deviation:** σ_x
 - *Weighing the deviation of the mean by the magnitude of the deviation*
- Demo on Github

Outline

- 1 Teacher and Student Introductions
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?
- 3 Course Outline
- 4 Dealing with Data
- 5 Manipulating Data in Math and Code
- 6 Some Statistics
- 7 Lab: Generating a Sine-Wave
- 8 Some Probability
- 9 Lab: Generating a Noisy Sine-Wave
- 10 Artificial Intelligence and Machine Learning
- 11 Why the Hype?

Lab: Generating a Sine-Wave

- Use the concepts to far to generate and plot a sine-wave in Python
- Go to the course Github
- Open the Day1 folder
- Open the Day1.ipynb notebook in Google Colab

Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Random Variables

- A **random variable** is a number whose value depends on the outcome of a random phenomenon

Random Variables

- A **random variable** is a number whose value depends on the outcome of a random phenomenon
 - Ex: a coin flip, noise in a signal

Random Variables

- A **random variable** is a number whose value depends on the outcome of a random phenomenon
 - Ex: a coin flip, noise in a signal
- We use random variables to model noise in data

Random Variables

- A **random variable** is a number whose value depends on the outcome of a random phenomenon
 - Ex: a coin flip, noise in a signal
- We use random variables to model noise in data
- A random variable is characterized by its **probability distribution**:

Random Variables

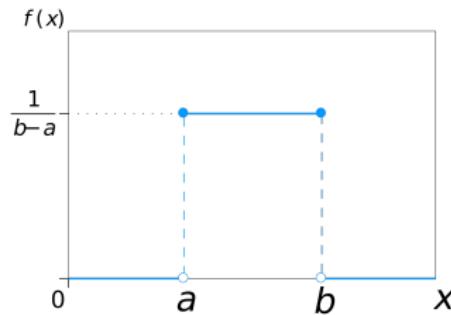
- A **random variable** is a number whose value depends on the outcome of a random phenomenon
 - Ex: a coin flip, noise in a signal
- We use random variables to model noise in data
- A random variable is characterized by its **probability distribution**:
 - a function that describes the “chance” or “likelihood” that an events will occur

Random Variables

- A **random variable** is a number whose value depends on the outcome of a random phenomenon
 - Ex: a coin flip, noise in a signal
- We use random variables to model noise in data
- A random variable is characterized by its **probability distribution**:
 - a function that describes the “chance” or “likelihood” that an events will occur
 - Ex: $P(\text{Coin flip} = \text{Heads}) = 0.5$

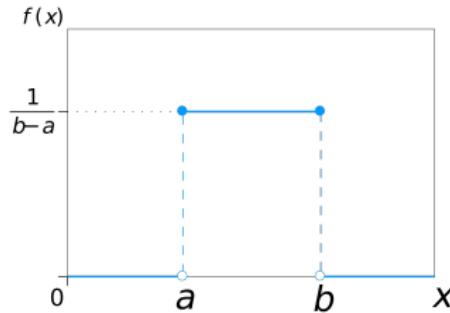
Uniform Distribution

- Uniform Distribution: $X \sim U(a, b)$



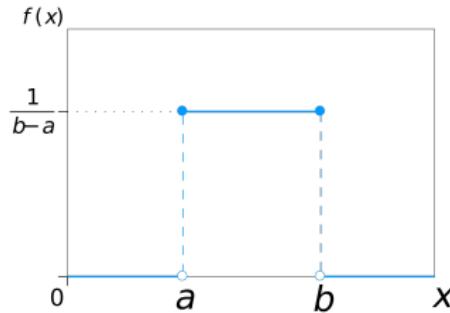
Uniform Distribution

- Uniform Distribution: $X \sim U(a, b)$
- The number, x , from the random variable X , is equally likely to fall anywhere on the interval $[a, b]$



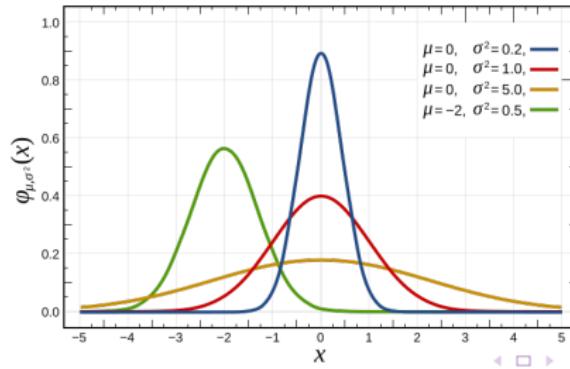
Uniform Distribution

- Uniform Distribution: $X \sim U(a, b)$
- The number, x , from the random variable X , is equally likely to fall anywhere on the interval $[a, b]$
- What is the mean of the distribution?



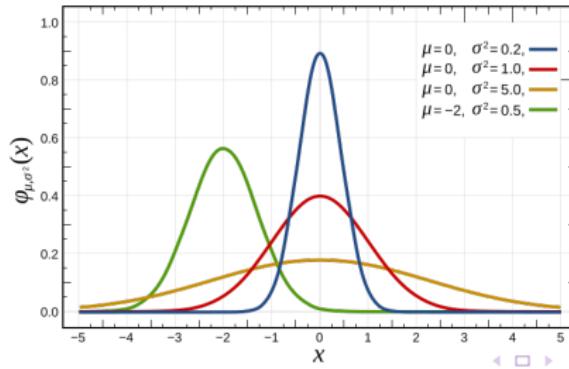
Gaussian Distribution (Normal Distribution)

- Gaussian Distribution: $X \sim N(\mu, \sigma^2)$



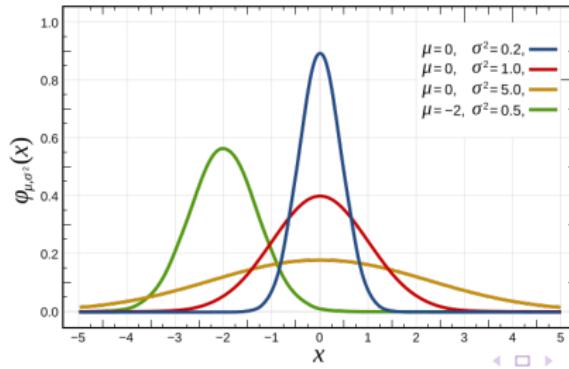
Gaussian Distribution (Normal Distribution)

- Gaussian Distribution: $X \sim N(\mu, \sigma^2)$
- Formula: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



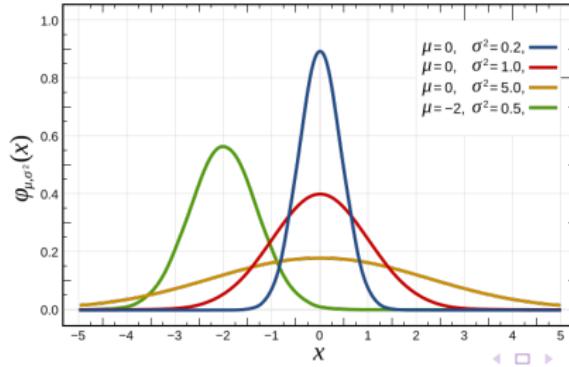
Gaussian Distribution (Normal Distribution)

- Gaussian Distribution: $X \sim N(\mu, \sigma^2)$
- Formula: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Determined by two parameters: mean μ , variance σ^2



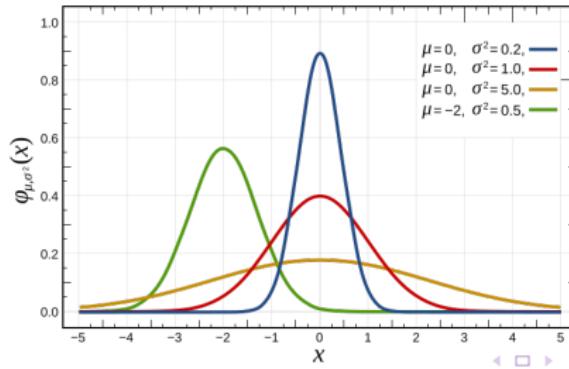
Gaussian Distribution (Normal Distribution)

- Gaussian Distribution: $X \sim N(\mu, \sigma^2)$
- Formula: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Determined by two parameters: mean μ , variance σ^2
- Often used to model noise in measurements



Gaussian Distribution (Normal Distribution)

- Gaussian Distribution: $X \sim N(\mu, \sigma^2)$
- Formula: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Determined by two parameters: mean μ , variance σ^2
- Often used to model noise in measurements
- **Standard Normal Distribution:** $N(0, 1)$



Outline

- 1 Teacher and Student Introductions**
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?**
- 3 Course Outline**
- 4 Dealing with Data**
- 5 Manipulating Data in Math and Code**
- 6 Some Statistics**
- 7 Lab: Generating a Sine-Wave**
- 8 Some Probability**
- 9 Lab: Generating a Noisy Sine-Wave**
- 10 Artificial Intelligence and Machine Learning**
- 11 Why the Hype?**

Lab: Generating a Noisy Sine-Wave

- Use the knowledge gained so far to generate and plot a noisy sine-wave in Python
- Again, use Numpy and Matplotlib
- Go to the Github, Day1/Day1.ipynb for the Colab template

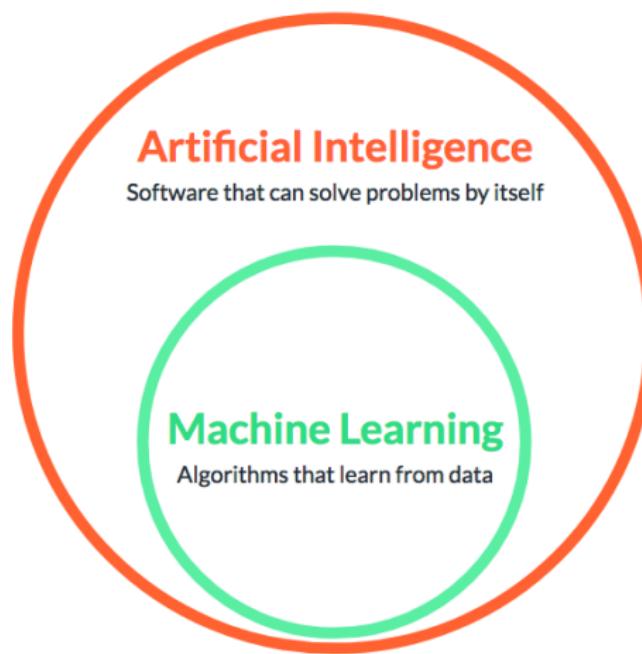
Outline

- 1 Teacher and Student Introductions
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?
- 3 Course Outline
- 4 Dealing with Data
- 5 Manipulating Data in Math and Code
- 6 Some Statistics
- 7 Lab: Generating a Sine-Wave
- 8 Some Probability
- 9 Lab: Generating a Noisy Sine-Wave
- 10 Artificial Intelligence and Machine Learning
- 11 Why the Hype?

Artificial Intelligence

- Search
- Reasoning and Problem Solving
- Knowledge Representation
- Planning
- Learning
- Perception
- Natural Language Processing
- Motion and Manipulation
- Social and General Intelligence

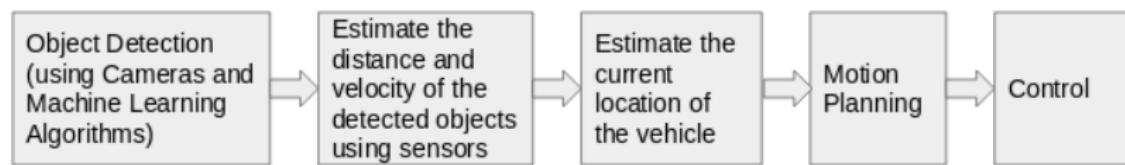
Machine Learning



Autonomous vs. Automated



Autonomous Example: Driver-less Cars



Outline

- 1 Teacher and Student Introductions
 - Teacher Introductions
 - Student Introductions
- 2 What is Machine Learning?
- 3 Course Outline
- 4 Dealing with Data
- 5 Manipulating Data in Math and Code
- 6 Some Statistics
- 7 Lab: Generating a Sine-Wave
- 8 Some Probability
- 9 Lab: Generating a Noisy Sine-Wave
- 10 Artificial Intelligence and Machine Learning
- 11 Why the Hype?

Why is Machine Learning so Prevalent?

- Database mining
- Medical records
- Computational biology
- Engineering
- Recommender systems
- Understand human brain

Why Now?

- Big Data

- Massive storage. Large data centers
- Massive connectivity
- Sources of data from internet and elsewhere

- Computational advances

- Distributed machines, clusters
- GPUs and hardware

Learning Objectives

- What is Machine Learning?
- What is Regression?
- What is Classification?
- Why is Machine Learning gaining so much importance?
- Where do I find data, how is it represented?
- How do we use Python to represent, manipulate, and visualize data?
- How do we use vectors and matrices to manipulate data?
- How do computers understand continuous data? How do we visualize it?
- How do we use random numbers in programming and modelling noise?

Thank You!

- Next Class: Linear Regression
- The real machine learning will begin!