Day 2: Linear Regression

Summer STEM: Machine Learning

Department of Electrical and Computer Engineering
NYU Tandon School of Engineering
Brooklyn, New York

Outline

1. Leftovers from Day 1
2. Introduction to Machine Leaning
3. Statistics Basics
4. Linear Regression

Exercises: Matrix Multiplication

$$(AB)_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj} \qquad (A^T)_{ij} = A_{ji}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 13 \end{bmatrix}$$

- $X = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}$ $Y = \begin{bmatrix} 3 & 1 \\ 0 & -1 \\ -2 & 3 \end{bmatrix}$ $Z = \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix}$
- Calculate XY, YX, $Z^T Y$

$$Z^T = \begin{bmatrix} 1 & 4 & 6 \end{bmatrix}$$

$$Z^T Y = \begin{bmatrix} 1 & 4 & 6 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 3 + (-12) & 1 - 4 + 18 \end{bmatrix} = \begin{bmatrix} -9 & 15 \end{bmatrix}$$

Matrix Inverse

A m by n matrix

$A^{-1}A = AA^{-1}$ m has to be equal to n

- Analogy: Reciprocal of a number $\frac{1}{a}a = 1$
- Matrix inverse only defined for square matrix (# rows = # cols)
- $A^{-1}A = AA^{-1} = I$. I is called the identity matrix, whose $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & & 1 \end{bmatrix}$ diagonal elements are 1 and other elements are 0.
- Hard to compute by hand, but for 2 by 2 matrix, it is

Ex: 2 by 2 identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$[a]^{-1} = [\tfrac{1}{a}]$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\underline{ad - bc}} \underline{\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}$$

- Even for a square matrix, the matrix inverse does not always exist. Can you tell when that is the case for 2 by 2 matrices based on the formula given above?

$ad - bc = 0$

NYU TANDON SCHOOL OF ENGINEERING

Matrix Inverse

When is matrix inverse useful? We can use it to solve systems of linear equations!

- Consider the following equations

$$\begin{cases} x + 2y = 5 \\ 3x + 5y = 13 \end{cases}$$

$$I u = u$$

$$\text{Ex} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} u_1 + 0 \\ 0 + u_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

- In matrix-vector form

$$\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 13 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}}_{I} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 13 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 13 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}^{-1} = \frac{1}{5 \times 1 - 2 \times 3} \begin{bmatrix} 5 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$$

Demo and Exercises: NumPy

Open `demo_vectors_matrices.ipynb`

- Your task: use NumPy functions to compute the exercises we did earlier this morning. Compare the results.

Demo: Plotting Functions

- Generate and plot the following functions in Python:
  - Scatter plot of points: (0,1), (2,3), (5,2), (4,1)
  - Straight Line: $y = mx + b$
  - Sine-wave $y = sin(x)$
  - Polynomial e.g. $y = x^3 + 2$
  - Exponential e.g. $y = e^{-2x}$
  - Choose a function of your own
- Use Wikipedia and Numpy Documentation to search for mathematical formulas and python functions

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

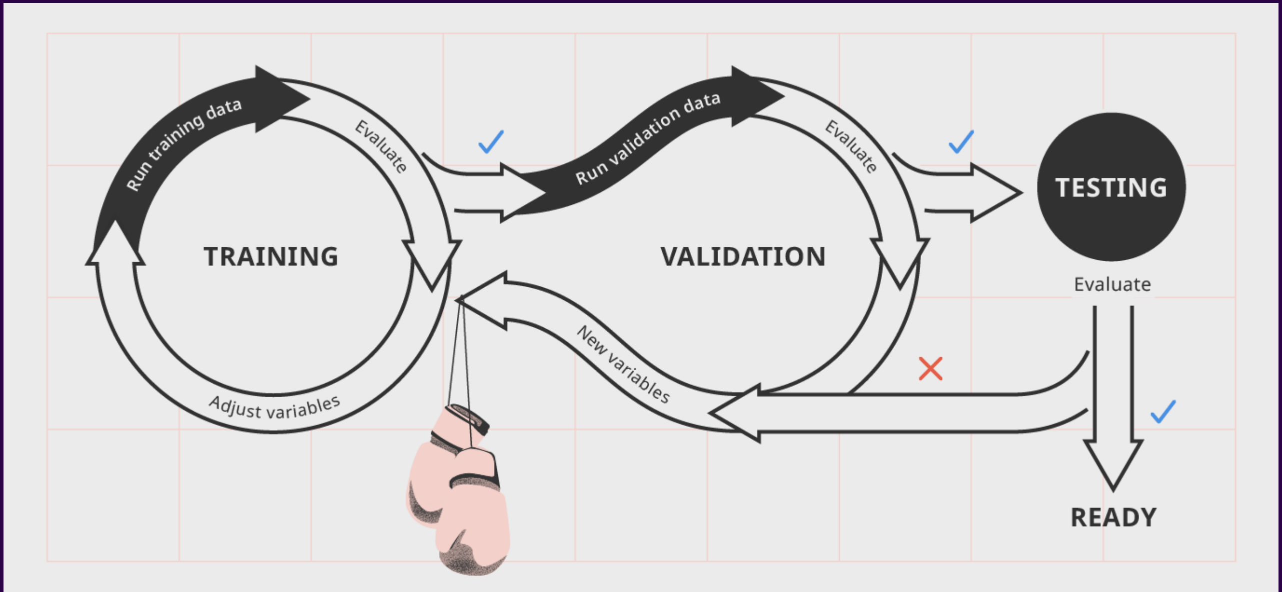
NYU TANDON SCHOOL OF ENGINEERING

Outline

What is Machine Learning

- Recognize patterns from data
- Make predictions based on the learnt patterns
- A very effective tool where human expertise is not available

Machine Learning Pipeline



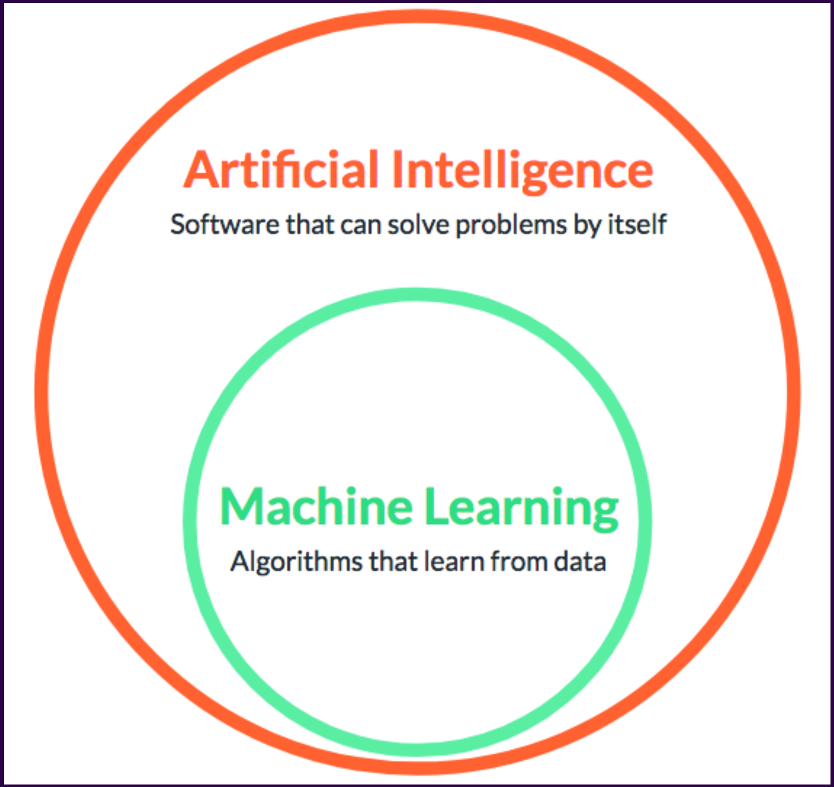
https://lionbridge.ai/articles/how-does-ai-training-work/

NYU TANDON SCHOOL OF ENGINEERING

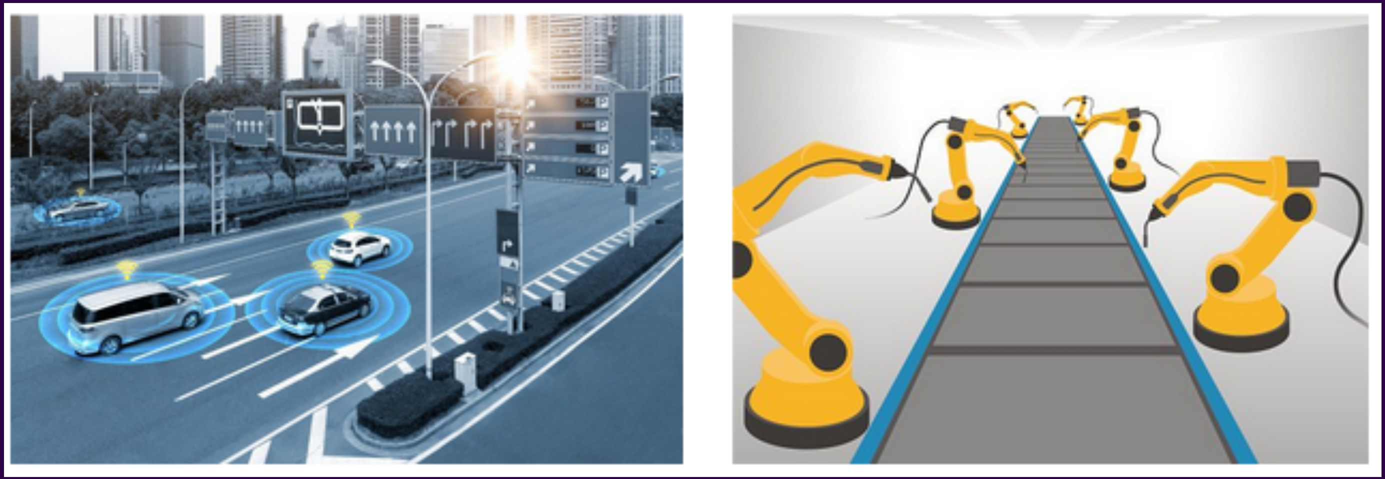
Artificial Intelligence

- 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: Self-driving car

- Waymo Video

---

https://www.tesmanian.com/blogs/tesmanian-blog/tesla-autopilot-full-self-driving-fsd-improvements-video

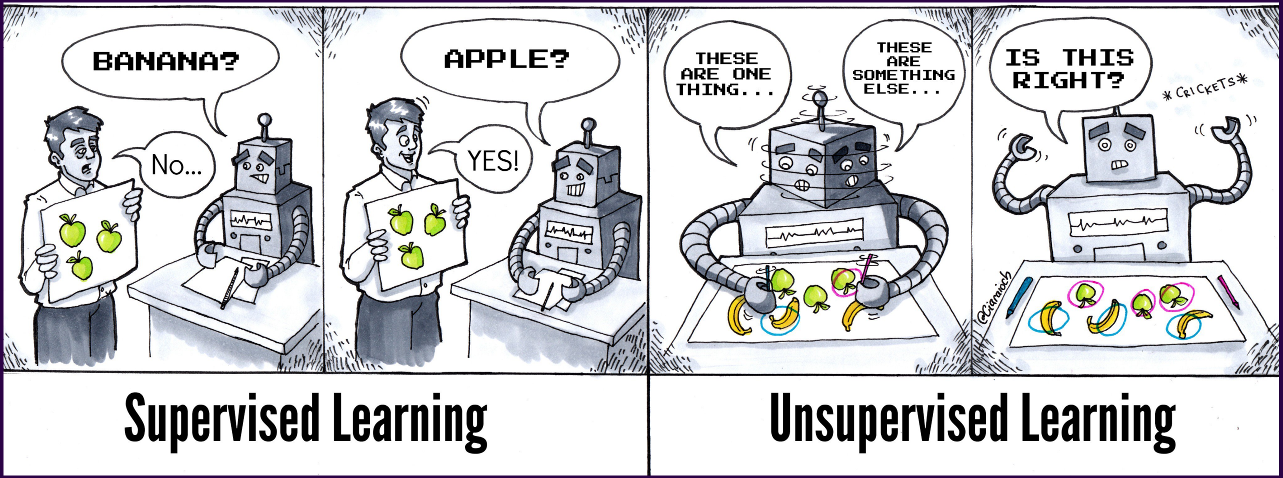
Why is Machine Learning so Prevalent?

- Database mining
- Medical records
- Computational biology
- Engineering
- Recommendation systems
- Understanding the 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

Supervised Vs. Unsupervised Learning

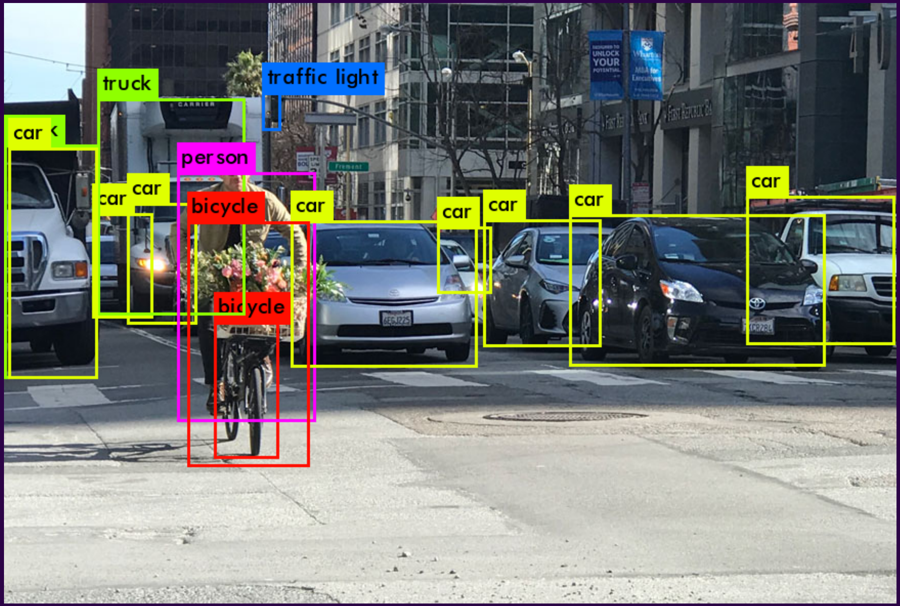


twitter.com/athena_schools/status/1063013435779223553/photo/1

Supervised Vs. Unsupervised Learning

- The main difference between supervised and unsupervised learning is the existence of a supervisor, which in many cases is in the form of a data label.
- The label of the data is what we want the machine learning algorithm to predict.

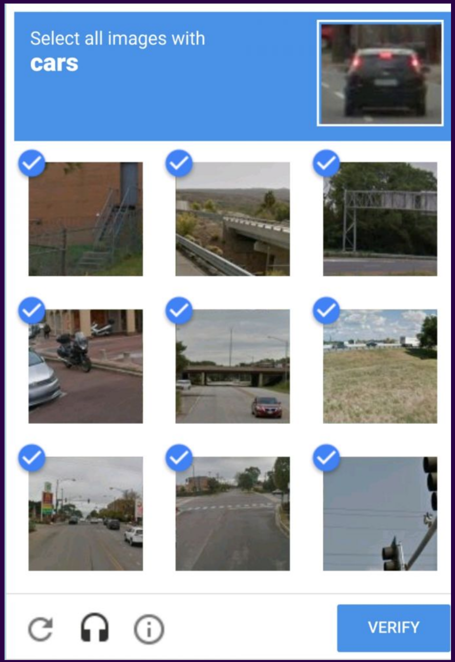
Labelled Data



- YOLO v2

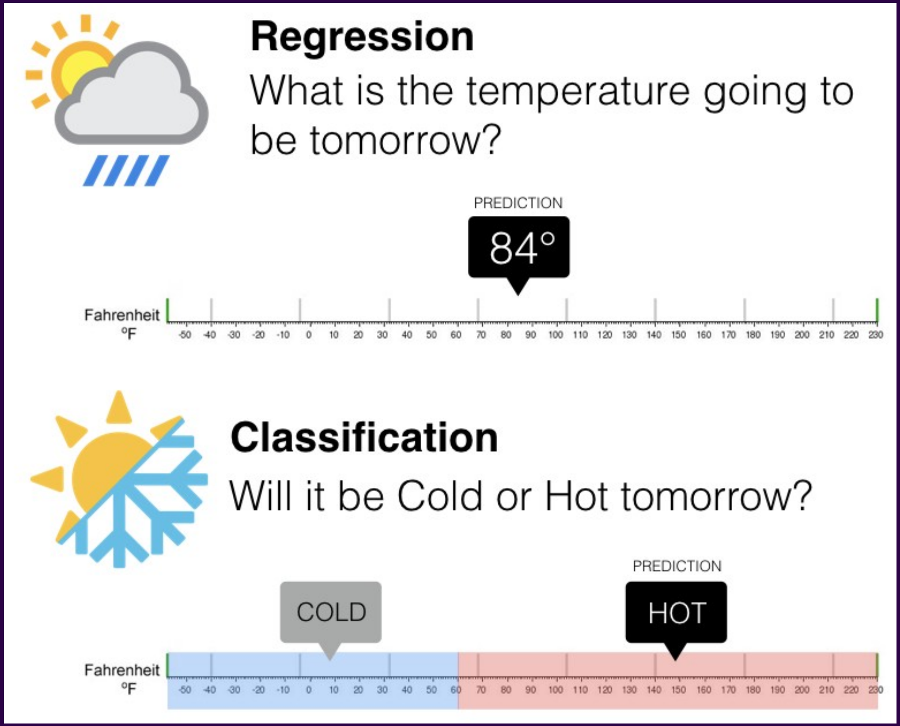
https://towardsdatascience.com/yolo-you-only-look-once-17f9280a47b0

How labels are generated



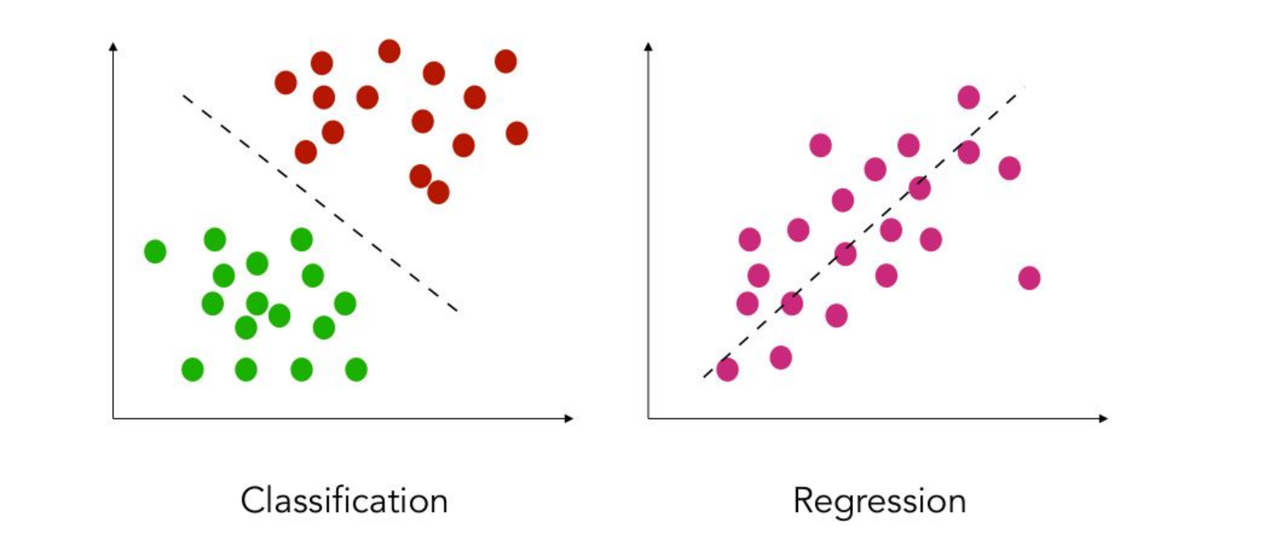
https://devrant.com/rants/1758134/select-all-images-with-cars-i-did-and-its-not-correct-why-not

Classification Vs. Regression



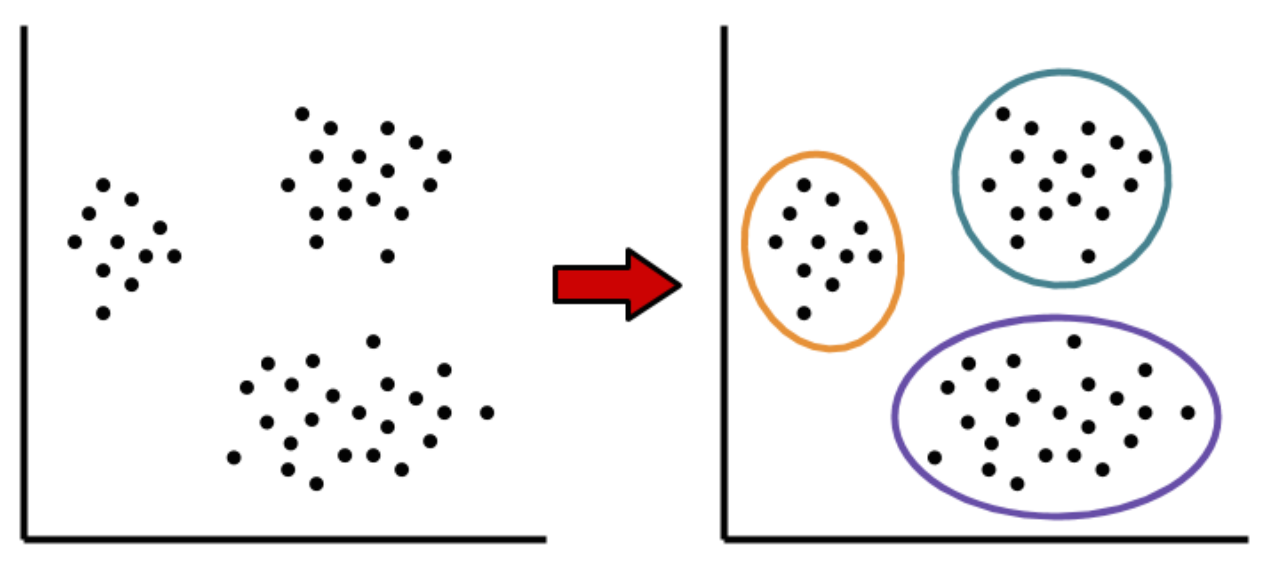
https://www.pinterest.com/pin/672232681855858622/?lp=true

Classification Vs. Regression



https://maelfabien.github.io/machinelearning/ml_base/

Unsupervised Learning



source: the dish on science

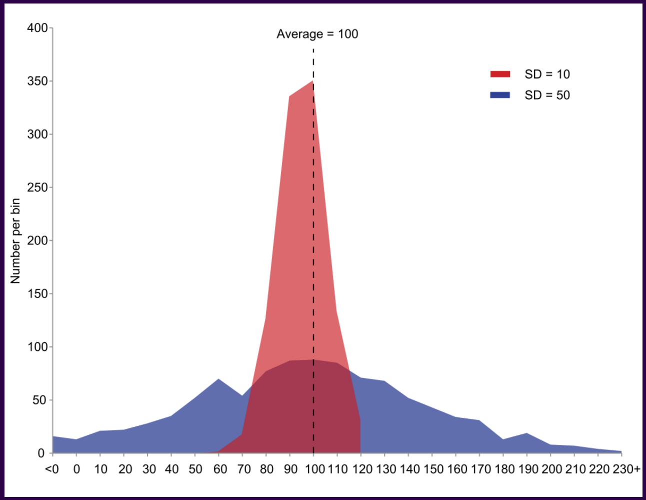
Outline

Basic Concepts

- **Mean** (average value): $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- **Variance** describes the spread of the data with respect to the mean.
- **Covariance** describes the relationship between two variables.

Variance

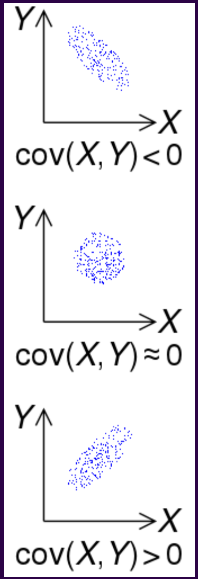
- Variance: $\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$



https://en.wikipedia.org/wiki/Variance

Covariance

- Covariance: $\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$



https://en.wikipedia.org/wiki/Covariance

NYU TANDON SCHOOL OF ENGINEERING

Outline

Supervised learning in a nutshell

Given the dataset (x_i, y_i) for $i = 1, 2, \ldots, N$, find a function $f(x)$ (model) so that it can predict the label $\hat{y}$ for some input x, even if it is not in the dataset, i.e. $\hat{y} = f(x)$.

Many possible models

- $f(x) = w_1 x + w_0$.
- $f(x) = w_2 x^2 + w_1 x + w_0$.
- $f(x) = \frac{1}{e^{-(w_1 x + w_0)} + 1}$.
- The numbers w_0, w_1 and w_2 are called model parameters.
- We often write the model as $f(x; \mathbf{w})$, stacking all parameters to a vector $\mathbf{w}$.

NYU TANDON SCHOOL OF ENGINEERING

How would you fit a line?

Can you find a line that passes through $(0, 0)$ and $(1, 1)$?

How would you fit a quadratic curve?

Can you find a quadratic curve that passes through $(0, 0)$, $(1, 1)$ and $(-1, 1)$?

What model do we use for this dataset?

- Open `demo_boston_housing_one_variable.ipynb`
- Can you find a line that go through ALL of the data points? Why?

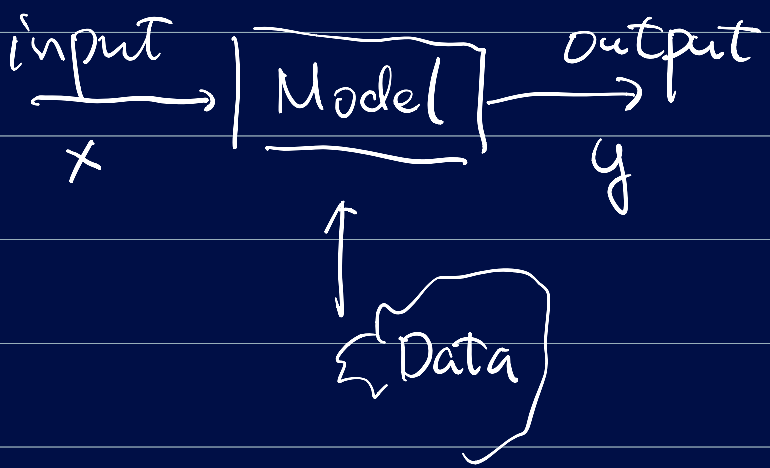
Is Your Model a Good Fit?

- How would you determine if your model is a good fit or not?
  - How will you determine this?
  - Is there a quantitative way?

Error Functions

- An **error function** quantifies the discrepancy between your model and the data.
  - They are non-negative, and go to zero as the model gets better.
- Common Error Functions:
  - Mean Squared Error: $MSE = \frac{1}{N}\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2$
  - Mean Absolute Error: $MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$
- In later units, we will refer to these as **cost functions** or **loss functions**.
- Compute MSE on your model
- How do we interpret MSE? MAE?

Machine Learning



Ex: x : movie genre

y : IMDB score

Dataset: $\underbrace{(x_i, y_i)}_{\text{a sample}}$ $i = 1, 2, \dots N$

We also call x features, y labels

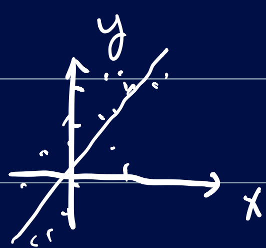
Task: to find a model $f(x)$ to predict

$$\hat{y} = f(x)$$

($\hat{y}$: "y hat", denotes the prediction given by the model)

Ex: $\hat{y} = f(x) = \underset{\sim}{w_1} x + \underset{\sim}{w_0}$

"linear model"



$w_0 = 0$

$w_1 = 2$

w_0, w_1 : parameters of the model

Cost functions : measure how well the model fits the data

MSE : $\frac{1}{N} \sum_{i=1}^{N} \| y_i - \hat{y}_i \|^2$

y_i : the labels in the dataset
"ground truth" label

$\hat{y}_i = f(x_i)$ predictions

Compute MSE in NumPy

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad \hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix} \quad \text{MSE}: \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$Y - \hat{Y} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_N - \hat{y}_N \end{bmatrix}$$

In NumPy, $z = (Y - \hat{Y}) ** 2$, this will square every element

$$\begin{bmatrix} (y_1 - \hat{y}_1)^2 \\ \vdots \\ (y_N - \hat{y}_N)^2 \end{bmatrix}$$

In our exercise $N = 506$

np.mean(z) will give $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$

(For a vector z with N elements

np.mean(z) gives $\frac{1}{N}\sum_{i=1}^{N} z_i$)