Foundation of ML

Week 1

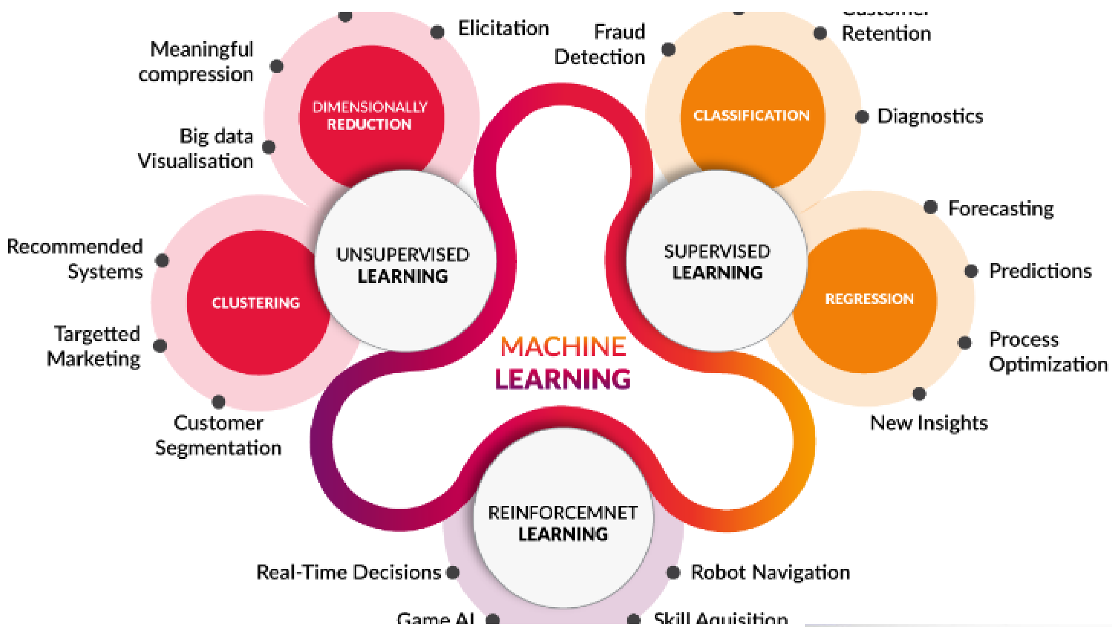
List of contents

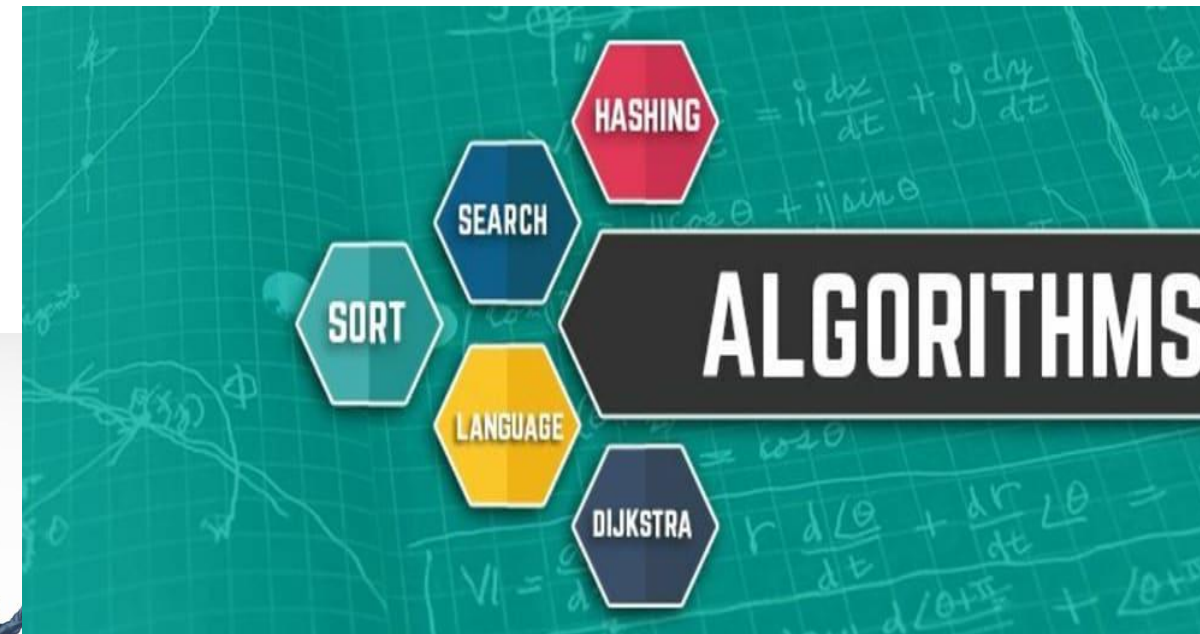
Introductory overview of:

- Real-world applications of Machine Learning (ML)
- Definitions of ML
- Types of ML
  - Unsupervised learning
  - Supervised learning
    - Linear models
    - Nonlinear models
- Model assessment and selection
- Vector, Matrix and operations

What is learning?

- See
- Listen
- Study
- Experience
- Do
- Taste
- Touch

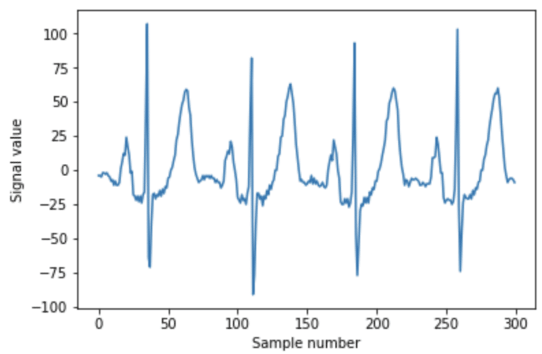


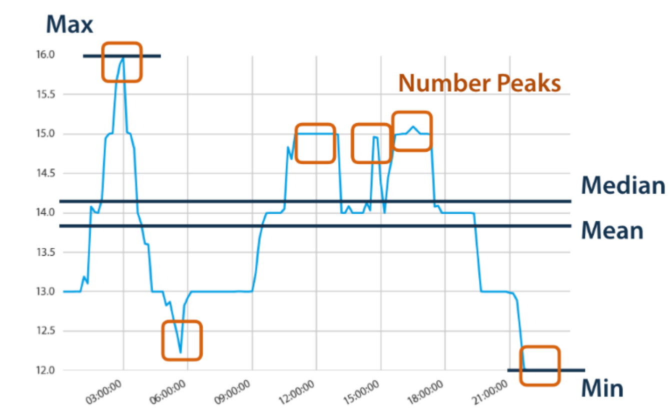


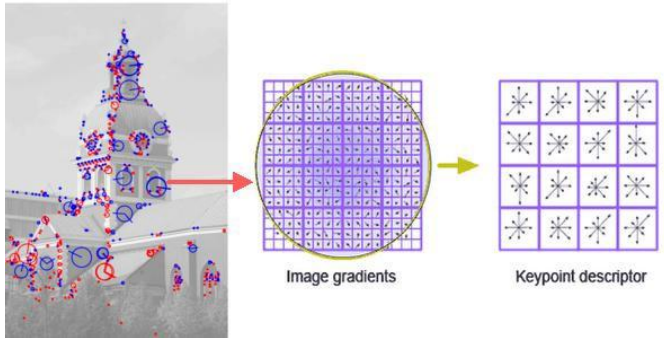
What is machine learning?

Steps in ML

- Data Collection
- Clean and Prepare data
- Train or Build Model
- Evaluate the model
- Improve/Deploy

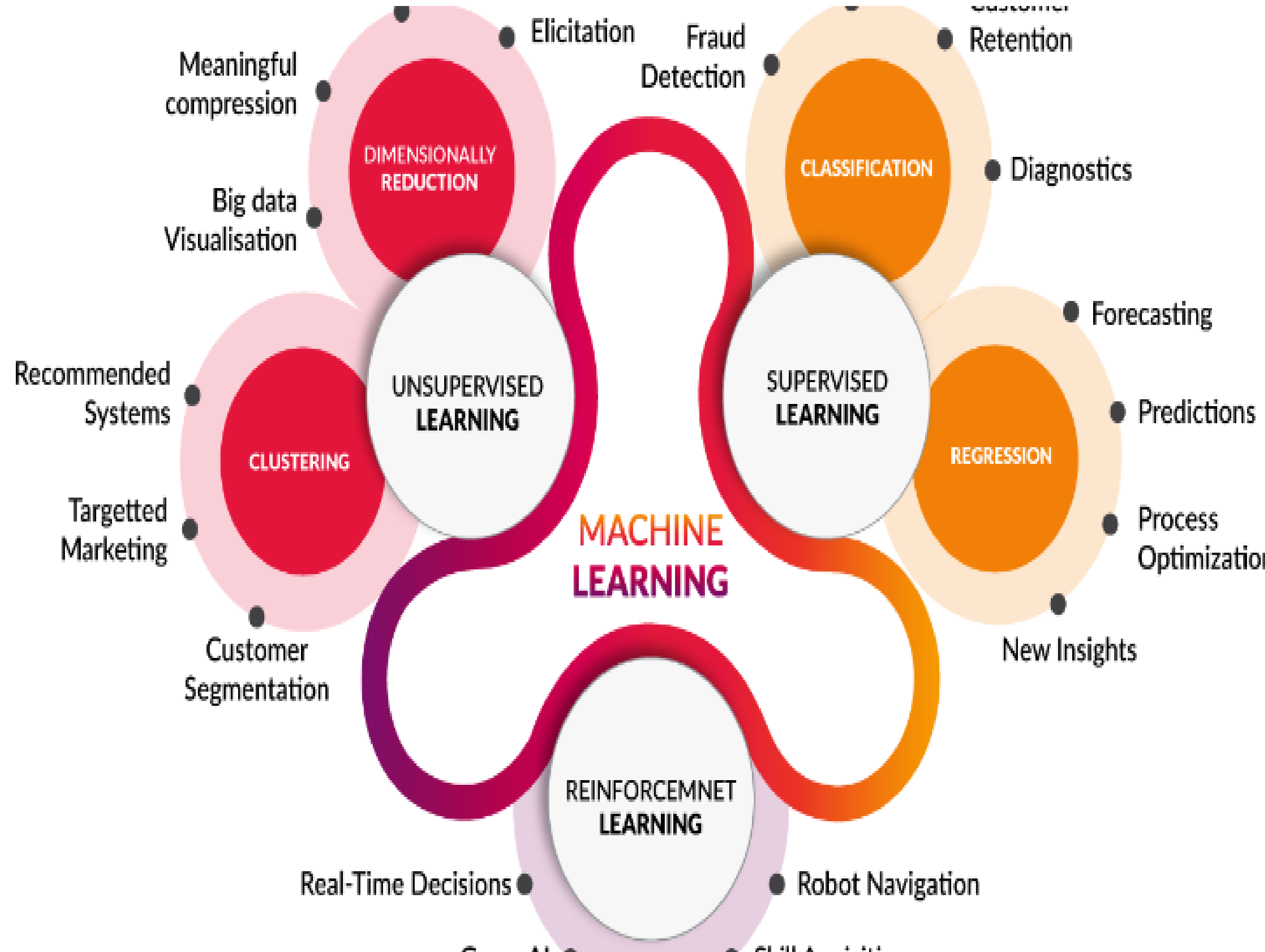






Understanding Data

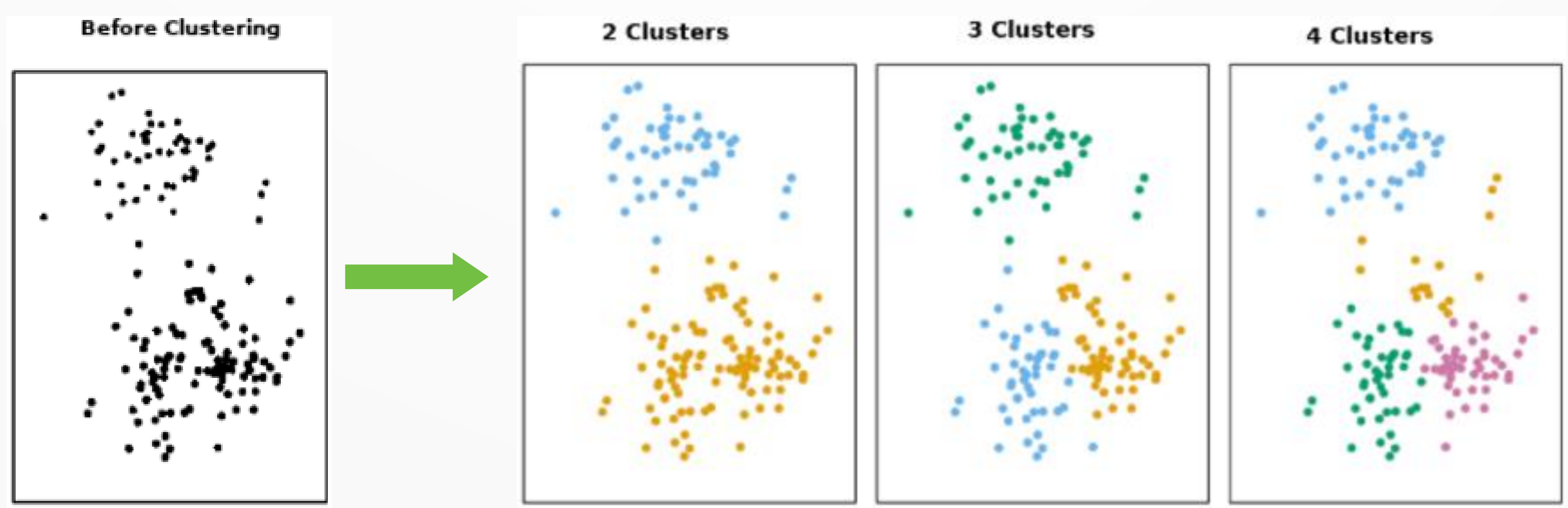
Types of Machine Learning Algorithm



Unsupervised algorithm

How do you find the underlying structure of a dataset which is unlabelled?

- **Clustering** (similarity-based)
  - the process of grouping similar points together
  - gives insight into underlying patterns of different groups



ML Type - Unsupervised learning...

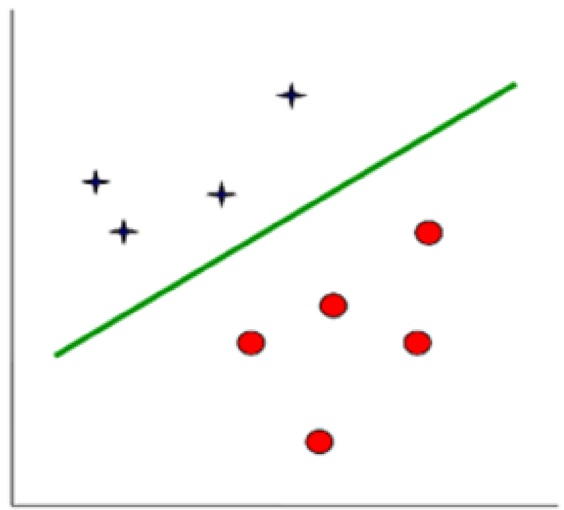
- Common examples -
  - Data understanding and visualization
  - Anomaly detection
  - Information retrieval
  - Data compression (reduction)

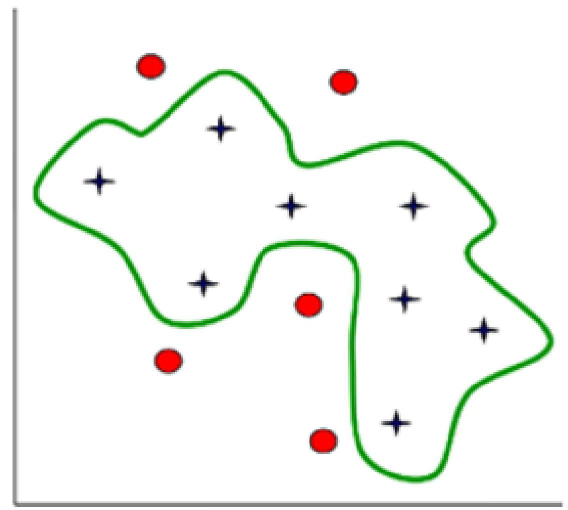
ML Type - Supervised learning

- "Learn a function (model) from data to relate the inputs with outputs."
- In supervised learning, the training data includes output information (labels/targets)
- **Target function**: $f:X \rightarrow Y$
- **Examples**: It is in the form of (x,y), denoted as $(x1,y1)$, ..., (xn,yn)
- **Hypothesis** $g:X \rightarrow Y$ such that $g(x)=f(x)$
  - x = set of attribute values
  - y = discrete label (*classification*), real valued number (*regression*)

ML Type - Supervised learning..

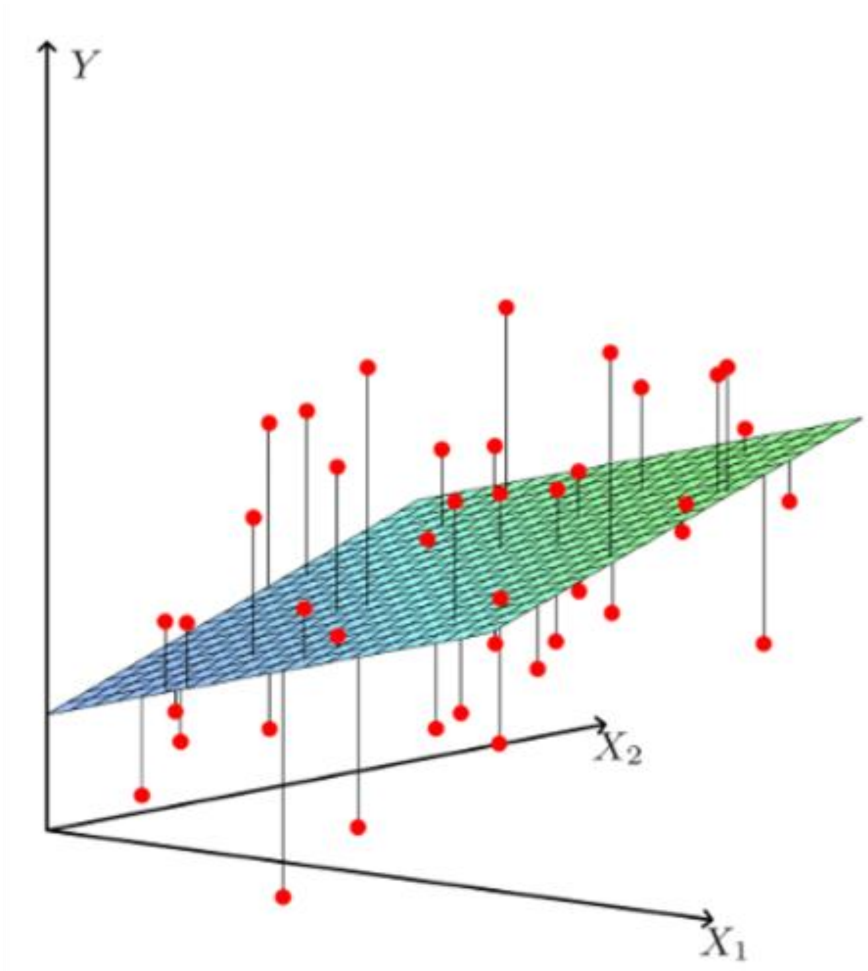
- Classification problem
  - with two classes, decision boundaries are a hyper-surface that partitions data space into two sets
  - each of these sets represents one of the classes
  - linear vs non-linear decision boundary



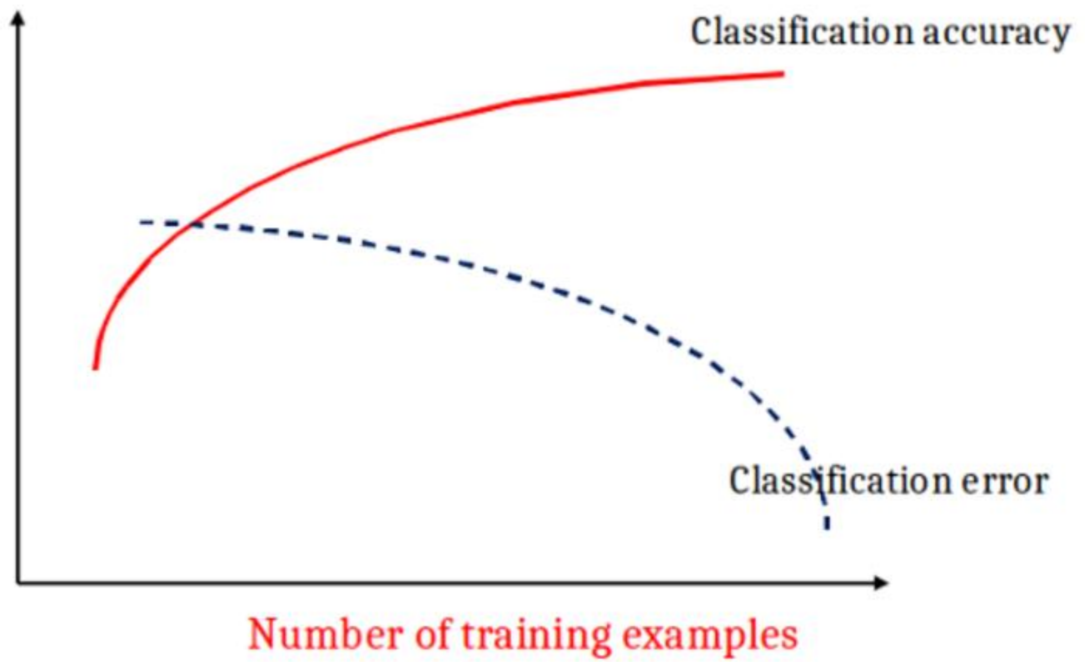


ML Type - Supervised learning...

- Regression problem
  - to examine the relationship between response variables and one or more predictor variables
  - examination can result in a hyperplane, representing the regression analysis
  - regression problem in 2 dimensions



Model assessment and adjustment



- Model evaluation
  - to determine if it will do a perfect job of predicting the labels on new and future test data
    - randomly split examples into a training set and test set
    - use training set to learn a model
    - evaluate the model using test set and a measurement (such as accuracy of prediction)
    - repeat for different random splits and average results
    - more training data, more accuracy

Model assessment and adjustment...

- Model selection
  - how to find the BEST model (hypothesis)?
  - There are often many knobs (parameters and hyper-parameters) that we can use to vary its fitness to the data
  - effective ways in which people approach this problem
    - look at averaged evaluation score on many random test sets
    - cross-validation (train using one set and test on the other, rotate them) etc.
  - be aware of *Over-fitting*

Revising knowledge of Linear Algebra

- Linear Algebra
  - Vector & their operations
  - Matrix & their operations

Vector

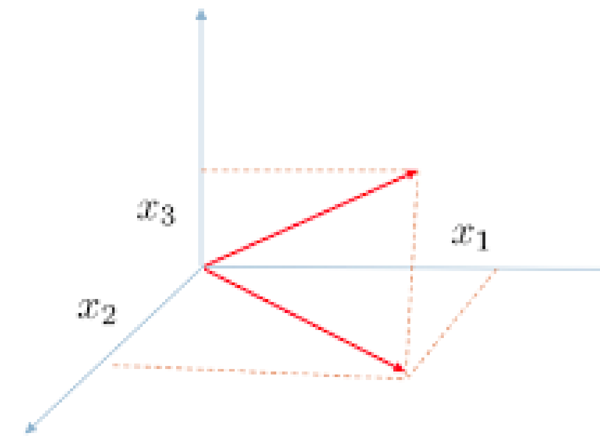
In machine learning algorithms:

A **data instance is represented by a vector**, more precisely, by a **feature vector.**

dimension

$$\mathbf{x} \in \mathbb{R}^n, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

x_i : i-th element



- Two simple vectors X and Y:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 & y_2 & \cdot & y_n \end{bmatrix}$$

Column vector

Row vector

Vector operations

- Three main operations in vectors -
  - Transpose
  - Addition
  - Inner product
- Transpose: $X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix}$

$$X^T = [x_1 \quad x_2 \quad .. \quad x_n]$$

Column vector

Row vector

Vector operations

Two vectors X and Y:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix} \qquad \mathrm{Y} = \begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ . \\ \mathrm{y}_n \end{bmatrix}$$

Addition operation:

$$X + Y = \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ . \\ x_n + y_n \end{bmatrix}$$

Inner product:

$$X^T Y = [x_1 y_1 + x_2 y_2 + .. + x_n y_n]$$

$$\mathrm{X} = [\mathrm{x}_1 \quad \mathrm{x}_2 \quad . \quad \mathrm{x}_n]$$

$$XY = [x_1 y_1 + x_2 y_2 + .. + x_n y_n]$$

Vector operations...

Magnitude of length of a vector:

$$length(X) = \sqrt{x_1^2 + x_2^2 + .. + x_n^2}$$

2-norm of a vector:

$$\|X\|_2 = \sqrt{x_1^2 + x_2^2 + .. + x_n^2}$$

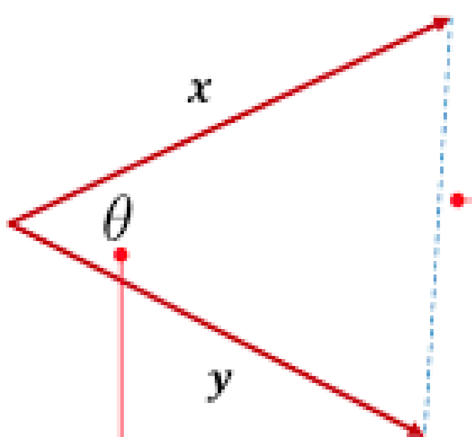
$$\|X\|_2 = (x_1^2 + x_2^2 + .. + x_n^2)^{\frac{1}{2}}$$

p-norm of a vector:

$$\|X\|_p = (|x_1|^p + |x_2|^p + .. + |x_n|^p)^{\frac{1}{p}}$$

Distances between vectors

Similarity



Euclidean distance

$$||x - y|| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

Cosine similarity

$$\cos(\theta) = \frac{x^T y}{||x||\,||y||} = \frac{x^T y}{\sqrt{x^T x}\sqrt{y^T y}}$$

Cosine distance is defined as

$$1 - \cos(\theta)$$

Matrix

Matrix has number of rows and columns:

$$A = \begin{array}{c} \begin{array}{cccc} \phantom{a_{11}} & \text{column vector} & & \phantom{a_{1n}} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

$$A = \begin{array}{cc} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} & \begin{array}{l} \\ \text{row vector} \\ \\ \\ \end{array} \end{array}$$

$$A = \begin{array}{c} \begin{array}{cccc} \phantom{a_{11}} & \phantom{a_{12}} & & \text{number of columns} \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\ \begin{array}{cccc} \text{number of rows} & \phantom{a_{12}} & & \phantom{a_{1n}} \end{array} \end{array}$$

number of columns

$$A \in \mathbb{R}^{m \times n}$$

number of rows

Matrix types

Rectangular and Square Matrices

- If a matrix A has size m×n such that m=n, then it is called a square matrix; otherwise, it is a rectangular matrix

$$\begin{bmatrix} 1 & 6 \\ 2 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 5 \\ 6 & 2 & 4 \end{bmatrix}$$

square matrix rectangular matrix

Type equation here.

Transpose of a Matrix

$$\begin{bmatrix} 1 & 6 & 7 \\ 2 & 3 & 8 \end{bmatrix} \overset{\text{transpose}}{\Longleftrightarrow} \begin{bmatrix} 1 & 2 \\ 6 & 3 \\ 7 & 8 \end{bmatrix}$$

$$A \leftrightarrow A^T$$

- Symmetric Matrices
  - a matrix is symmetric if it is equal to its transpose

$$\begin{bmatrix} 2 & 3 & 7 \\ 3 & 5 & 9 \\ 7 & 9 & 1 \end{bmatrix}$$

- Symmetric matrices are always square.

- Diagonal Matrix
  - A matrix A is called a diagonal matrix if A(i,j)=0 for all i≠j.
  - Diagonal matrix is always a square matrix.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

- Identity Matrix
  - A matrix *I* is called an identity matrix if it is a diagonal matrix and *I(i,i)*=1
  - denotes *nxn* identity matrix.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix operations

Matrix Addition/Subtraction (two matrices of same size)

$$X + Y = \begin{bmatrix} 2 & 4 \\ 3 & 1 \\ 8 & 5 \end{bmatrix} + \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 11 \\ 7 & 5 \\ 9 & 8 \end{bmatrix}$$

Scalar Multiplication or Division (multiply each element of matrix by the scalar)

$$3x \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 18 & 21 \\ 12 & 12 \\ 3 & 9 \end{bmatrix}$$

Element wise Matrix Multiplication (matrices have the same size)

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \\ 8 & 5 \end{bmatrix} \odot \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 28 \\ 12 & 4 \\ 8 & 15 \end{bmatrix}$$

Matrix operations

Matrix to Matrix Multiplication

- number of columns in the first matrix is equal to the number of rows in the second matrix
- Consider AB=C. Now C(i,j) is computed by dot product of A(i,:) and B(:,j)

$$\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 11 & 28 \\ 8 & 23 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 11 & 28 \\ 8 & 23 \end{bmatrix}$$

- Matrix multiplication is NOT commutative. Multiplication order matters. In general AB≠BA.

Matrix operations

- Inverse Matrix
  - Matrix A is called as inverse of matrix B, if and only if *BA=AB=I*.
  - Since AB=BA, both A and B need to be a square matrix
  - If A is inverse of B, we denote it as

$$A = B^{-1}$$

  - Inverse of a matrix A exists only if it’s determinant is nonzero

Summary

Data

Model

Decision

Thank You.