Chapter I
Analytic Geometry

PHOK Ponna, PHAUK Sokkhey and NEANG Pheak

Institute of Technology of Phnom Pech

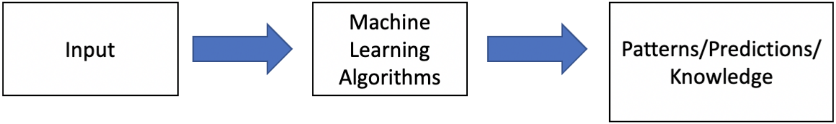
January 2023

Outline

Introduction to mathematics for Machine Learning

Definition

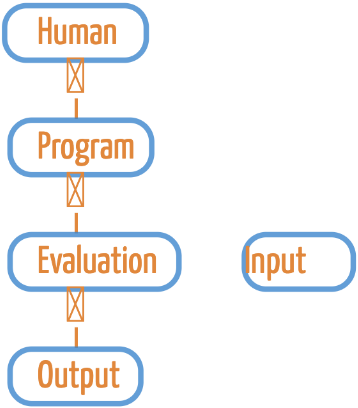
Machine learning is a field of computer science that gives computers the ability to learn without explicit programming, Arthur Samuel (1959).



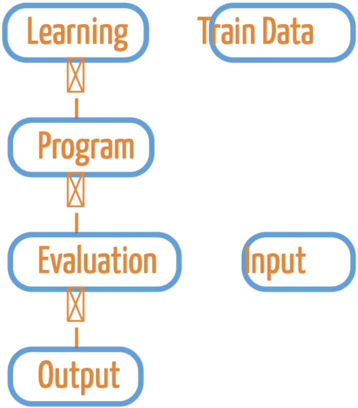
- The input needs to be relevant to given Algorithms.
- Choose Algorithms for optimal results.

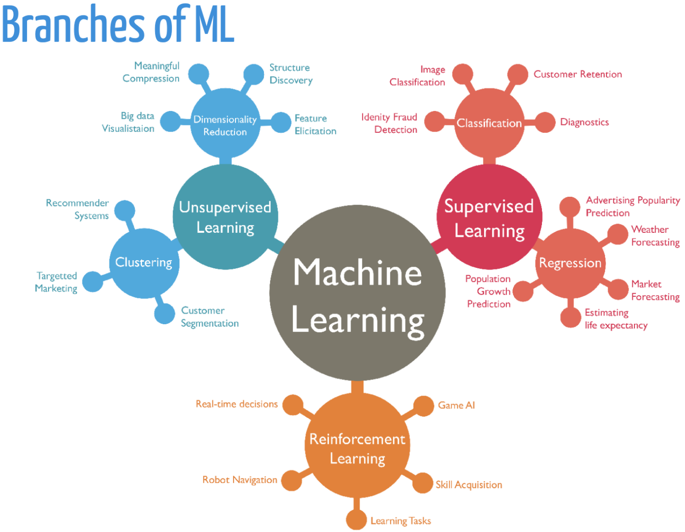
TP vs ML

- Traditional Programming:



- Machine Learning




Branches of ML
Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Data

- It can be: images, videos, voices, texts, time series, survey, ...
- Type: numerical, categorical,...
- General form:

Id	X_1	X_1	$\dots$	X_m
$\mathbf{x}_1$	x_1^1	x_1^2	$\dots$	x_1^m
$\mathbf{x}_2$	x_2^1	x_2^2	$\dots$	x_2^m
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{x}_n$	x_n^1	x_n^2	$\dots$	x_n^m

- Individual: $\mathbf{x}_i = (x_i^1, \dots, x_i^m)^T$, for $i = 1, \dots, n$.
- Variable: $X_j = (x_1^j, \dots, x_n^j)$, for $j = 1, \dots, m$.

First example

- **Iris**: `Hello world` data of **ML** (150 rows & 5 columns).

Sepal Lth	Sepal Wth	Petal Lth	Petal Wth	Species
5.1	3.5	1.4	0.2	setosa
6.9	3.1	4.9	1.5	versicolor
6.9	3.1	5.4	2.1	virginica
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
5.6	3	4.5	1.5	versicolor

  - Flower $\mathbf{x_1} = (5.1, 3.5, 1.4, 0.2, \texttt{setosa})$.
  - Variable $\textbf{Species} = (\texttt{setosa}, \texttt{versicolor}, \texttt{virginica}, \texttt{versicolor})$.

Types of Algorithms:

- Supervised learning, given a data set $\{(x_i, y_i)\}, \exists$ a relation $f : X \to Y$.

$$\begin{cases} \text{Given: Training set}\{(x_i, y_i)\}|i = 1, \cdots \mathbf{N}\}. \\ \text{Find : } \hat{f} : X \to Y, \text{a good approximation to } f \end{cases}$$

Remark

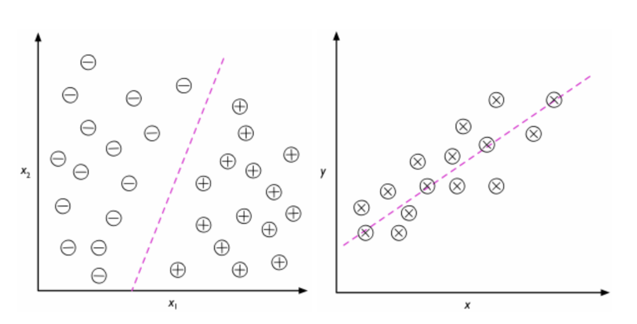
Very often, $X = \mathbf{R}^{\mathbf{d}}$,and

- $Y = \mathbf{R}$: Regression.
- Y is finite ($|Y|$ is not big): classification.

Introduction to Mathematics for Machine Learning

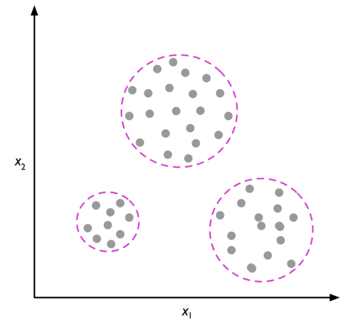
Example 1: Emails SPAM classification

Example 2: classification and regression



- Unsupervised learning: only inputs are supplied to the algorithms. There are no outputs. The algorithms simply study the inputs and put them as clusters or tell us how the input is associated with one another.

Example 3: Clustering

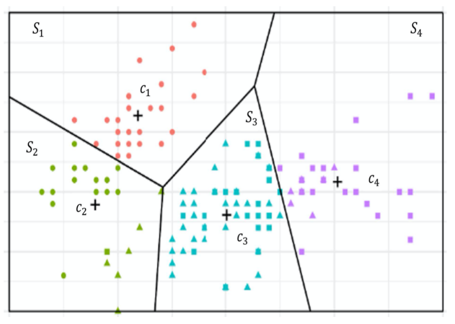


Clustering

Definition

clustering consists in partitioning a set of points from some metric space in such a way that points within the same group are close enough while points from different groups are distant. It belongs to the unsupervised learning branch of Machine learning (ML).

Example 4



Clustering in ML application

clustering shows up in many Machine learning applications, for example:

- Marketing: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records.
- biology: classification of plants and animals gave their feature.
- Insurance: Identifying groups of motor insurance policy holders with a high average claim cost; identifying frauds.
- Bookshops: book ordering(recommendation)
- City-planning: identifying groups of houses according to their type, value, and geographical location.
- Internet: document classification; clustering web log data to discover groups of similar access patterns; topic modeling $\cdots$

Introduction to Mathematics for Machine Learning

Mathematical foundation for Machine learning

- Linear Algebra Data Structures
- Tensor Operations
- Matrix Properties
- Eigenvector and Eigenvalue
- Matrix Operations for Machine Learning
- Calculus: Differentiation for optimization, Integration for computing the average
- Probability for modeling
- Statistical prediction and modeling

Scalars, Vectors, and Vector transpositions

Definition

Scalars: A scalar is a single number that has no dimension.

- denoted with regular typed 1 or 21.

Definition

Vectors: A vector is an array of numbers, that is, a single row or column of numbers.

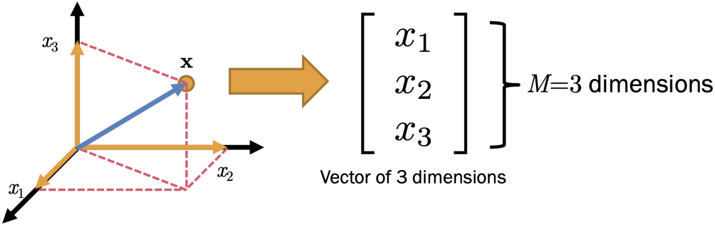
- denoted with bold small letters
- row vector: eg. $\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}$
- column vector (default): eg. $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$

Scalars, Vectors, and Vector tranpositions

Remark

Vector

- A point in the M-dimensional space
- Denoted by $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, etc.
- Usually treated as a data point



Vector Tranpositions

Definition

Given a vector $\mathbf{x} \in \mathbb{R}^{1\times n}$, $\mathbf{x} = \begin{bmatrix} x_1 & x_1 & \cdots & x_n \end{bmatrix}$, then the transpose of $\mathbf{x}$, denoted as $\mathbf{x}^T \in \mathbb{R}^{n\times 1}$, where $\mathbf{x}^T = \begin{bmatrix} x_1 \\ x_1 \\ \dots \\ x_n \end{bmatrix}$.

- Intuitionly: "the rows become columns, the columns become the rows"

Example

$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then $\mathbf{x}^T = \begin{bmatrix} 1 & 2 \end{bmatrix}$

,

Norms

Definition

A norm on a vector space V is a function

$$\begin{aligned} \|\cdot\| :\ & \mathbf{v} \to \mathbb{R}, \\ & \mathbf{x} \mapsto \|\cdot\|, \end{aligned}$$

which assigns each vector $\mathbf{x}$ its length $\|\mathbf{x}\| \in \mathbb{R}$,s.t for all $\lambda \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$ the following holds:

- $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- $\|\mathbf{x}\| \geq 0$ or $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$

Note: A norm of a vector $\|\mathbf{x}\|$ is informally a measure of the "length" of the vector.

Norms

Example 1

- l_1 norm and l_2 norm (The Euclidean Norm) on $\mathbb{R}^n$ are defined for $\mathbf{x} \in \mathbb{R}^n$ respectively as

$$\|\mathbf{x}\|_1 := \sum_{i=1}^{n} |\mathbf{x}_i|, \tag{1}$$

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^{n}} = \sqrt{\mathbf{x}^T \mathbf{x}} \tag{2}$$

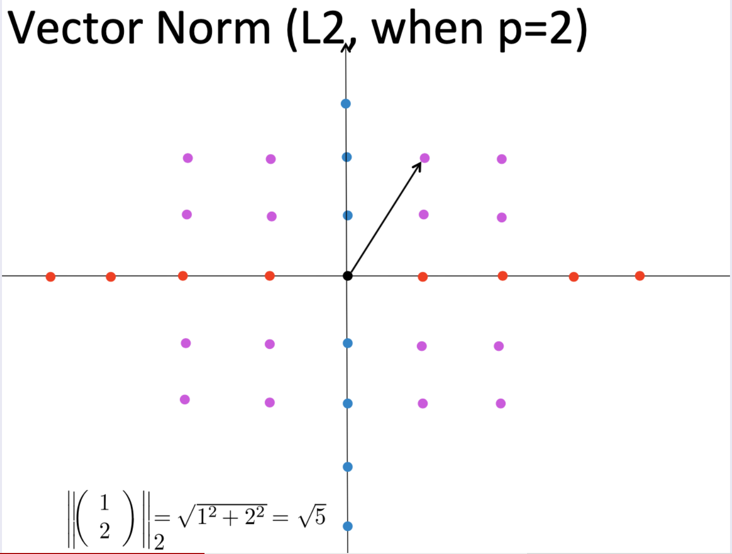
- Infinity norm (l_∞)

$$\|\mathbf{x}\|_\infty := \max\left(|\mathbf{x_i}|\right)$$

Note: $\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$, where $p \geq 1$.

Norms

Example 2



Dot Product

Definition

The Dot product is a way to multiply two equal-length vectors together. Conceptually, it is the sum of the products of the corresponding elements in the two vectors. Other names for the same operation include:

- Scalar product
- Inner product
- Projection product.

Remark

Given $\mathbf{x} = \begin{bmatrix} x_1 & x_1 & \cdots & x_n \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} y_1 & y_1 & \cdots & y_n \end{bmatrix}$, then the dot product is given by

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=n}^{n} x_i y_i = x_1 y_1 + \cdots x_n y_n = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta$$