

Machine Learning

▣ Classification of Machine Learning Algorithms →

⊕ Supervised

↳ Labeled data available

⊕ Unsupervised

↳ Data not Labeled

⊕ Reinforcement

↳ Reward based learning.

Classification of Supervised Learning

↓
Classification
Problem

↓
Regression
Problem

Algorithms to solve classification problem:

→ K-nearest neighbors (KNN)

→ Support Vector Machine (SVM)

→ Decision Tree

→ Random Forest.

KNN (K-Nearest Neighbors)

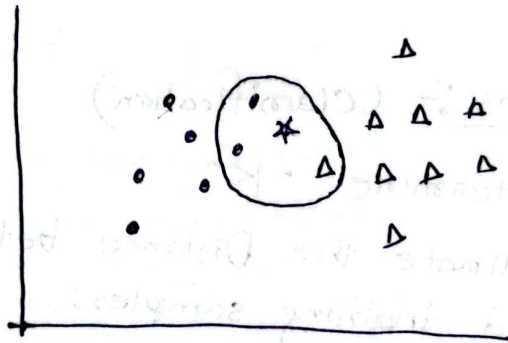
$$L_1 = |x_2 - x_1| + |y_2 - y_1|$$

$$L_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Hence,

L_1 = Manhattan Distance

L_2 = Euclidean Distance



- * Hence, Suppose, the value of K is 3
- * 2 of the nearest points are a circle
- * So, the query point belongs to the class circle.

KNN

* K-NN is a Supervised Machine learning Algorithm

* The Perception behind it

Given some training data and a new data point, we would assign the ~~new~~ new data based on the class of the training data it is nearest to.

The algorithms :- (Classification)

- (a) Determine "K".
- (b) Estimate the Distance between new data and training samples.
- (c) Sort the distance
- (d) Collect the classes of the top '3'
- (e) Choose the majority one.

▣ Let's see a Regression Task:

S.No	Height	Age	Weight
P ₁	6	40	60
P ₂	6.11	26	55
P ₃	5.9	30	56
P ₄	5.8	32	54
P ₅	5.3	33	75
P ₆	5.6	34	78
P ₇	5.5	35	80
P ₈	5.8	37	?

Step-1: The distance between P₈ and rest of training set are calculated using Euclidean metric.

- ① $\sqrt{(5.8-6)^2 + (37-40)^2} = 3.006$ (P₁)
- ② $\sqrt{(5.8-6.11)^2 + (37-26)^2} = 11.007$ (P₂)
- ③ $\sqrt{(5.8-5.9)^2 + (37-30)^2} = 7.0007$ (P₃)
- ④ $\sqrt{(5.8-5.8)^2 + (37-32)^2} = 5$ (P₄)
- ⑤ $\sqrt{(5.8-5.3)^2 + (37-33)^2} = 4.03$ (P₅)
- ⑥ $\sqrt{(5.8-5.6)^2 + (37-34)^2} = 3.006$ (P₆)
- ⑦ $\sqrt{(5.8-5.5)^2 + (37-35)^2} = 2.022$ (P₇)

Step-2: Sort the distance Top-3 nearest neighbours are (K=3) P₇, P₁ and P₆

Step-3: Predict P₈'s weight by taking the mean of the weights of P₇, P₁, P₆

$$\therefore P_8's \text{ weight} = \frac{60 + 78 + 80}{3} = 72.66 //$$

* Let's say, we have a document classification task. Let a document d_5 ^{x_1, x_2} $(3, 7)$ needs to be assigned a class. Below are four training documents belonging to class $'C_1'$ & $'C_2'$

	x_1	x_2	Euc Distance	class
d_1	7	7	$\sqrt{(7-3)^2 + (7-7)^2} = 4$	C_1
d_2	7	4	$\sqrt{(7-3)^2 + (7-4)^2} = 5$	C_2
d_3	3	4	$\sqrt{(3-3)^2 + (4-7)^2} = 3$	C_1
d_4	1	4	$\sqrt{(1-3)^2 + (4-7)^2} = 3.605$	C_1

Hence, Top 3 near documents & their ~~class~~ classes

are $\rightarrow d_3 \rightarrow C_1$

$d_4 \rightarrow C_1$

$d_1 \rightarrow C_1$

\therefore The class $'C_1'$ gets the majority vote and the new document d_5 is assigned to class $'C_1'$.

Practice Problem (KNN)

* Determine the BMI for the query point (150, 61)

Ans:

Here, the distance between the query point (150, 61) and rest of the training set calculating using Euclidian metric:

- ① $\sqrt{(150-150)^2 + (61-45)^2} = 16$ BMI Normal
- ② $\sqrt{(150-155)^2 + (61-40)^2} = 10.29$ O.W
- ③ $\sqrt{(150-145)^2 + (61-35)^2} = 26.47$ U.W
- ④ $\sqrt{(150-160)^2 + (61-55)^2} = 11.66$ Normal
- ⑤ $\sqrt{(150-140)^2 + (61-50)^2} = 14.86$ O.W
- ⑥ $\sqrt{(150-142)^2 + (61-30)^2} = 32.01$ U.W
- ⑦ $\sqrt{(150-152)^2 + (61-65)^2} = 4.47$ O.W
- ⑧ $\sqrt{(150-157)^2 + (61-42)^2} = 20.24$ Normal
- ⑨ $\sqrt{(150-147)^2 + (61-42)^2} = 19.23$ Normal
- ⑩ $\sqrt{(150-150)^2 + (61-50)^2} = 11$ O.W
- ⑪ $\sqrt{(150-145)^2 + (61-60)^2} = 5.09$ O.W
- ⑫ $\sqrt{(150-155)^2 + (61-40)^2} = 21.58$ Normal
- ⑬ $\sqrt{(150-152)^2 + (61-50)^2} = 11.18$ O.W
- ⑭ $\sqrt{(150-145)^2 + (61-70)^2} = 10.29$ U.W
- ⑮ $\sqrt{(150-140)^2 + (61-25)^2} = 37.36$ O.W
- ⑯ $\sqrt{(150-147)^2 + (61-29)^2} = 32.14$ U.W
- ⑰ $\sqrt{(150-149)^2 + (61-71)^2} = 10.04$ Normal U.W
- ⑱ $\sqrt{(150-157)^2 + (61-59)^2} = 4.28$ Normal

Height	Weight	BMI
150	45	Normal
155	40	Overweight
145	35	Underweight
160	55	Normal
140	50	Overweight
142	30	Underweight
152	65	Overweight
157	42	Normal U.W
147	42	Normal
150	50	Overweight
145	60	Normal O.W
155	40	Underweight
152	50	Underweight Normal
145	40	Overweight O.W
140	25	Underweight U.W
147	29	Underweight
149	41	Underweight
157	59	Normal
153	37	Underweight