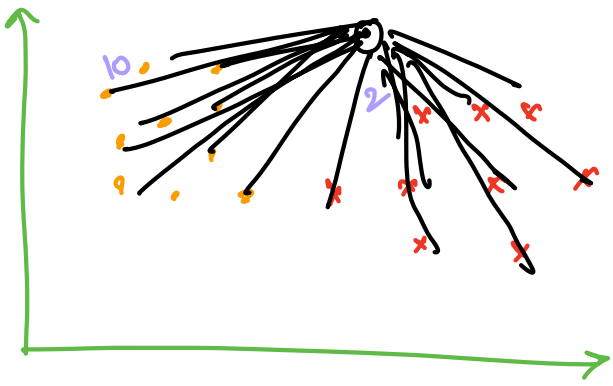
- Linear Reg. → Reg.
- Logistic Reg → Class. → weights (training data)

K Nearest Neighbours (KNN)

Training Time: O(1)

↳ Classification



(1) Distance b/w test data point & all the training data

(2) Distance sorted order arrange

- small 2
- large 10

(3) k given
↳ 3

(4) majority class test data point

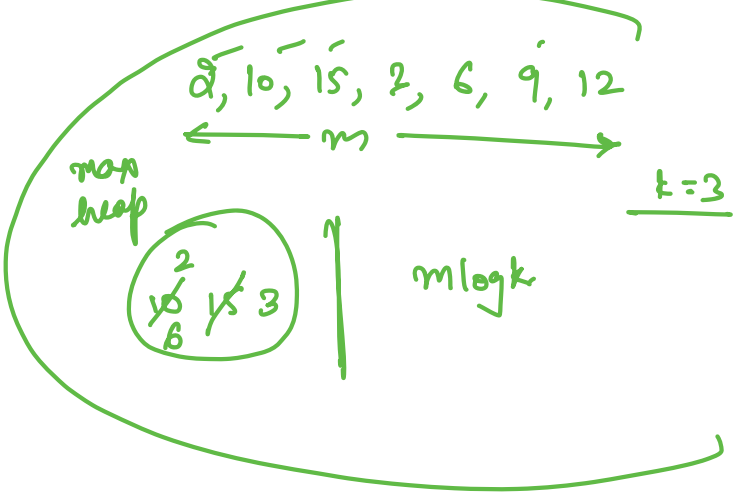
Training Time : O(1)

Test Time: m + $m\log m + k$ (heap : $m\log k$)

m → distance; $m\log m + k$ → sorting → topk

: $m + m \log k$

: $O(m \log k)$



Hoare Selection Algo

A factory is producing papers. The quality control unit applies two types of testing (durability test and strength test) to assess paper quality. The data for the same is given below:

Table III

S. No.	1	2	3	4	5	6	7	8
Durability	7	6	7	6	3	1	4	3
Strength	7	4	4	5	4	4	3	5
Quality	Good	Bad	Good	Good	Bad	Bad	Bad	Bad

~~In general, the factory produces 720 good quality papers out of 1000.~~
Use k-nearest neighbor (KNN) with k = 1, and 3 to predict the quality of a new paper (durability = 5, strength = 5). **[2+1] [CO3] [L3]**

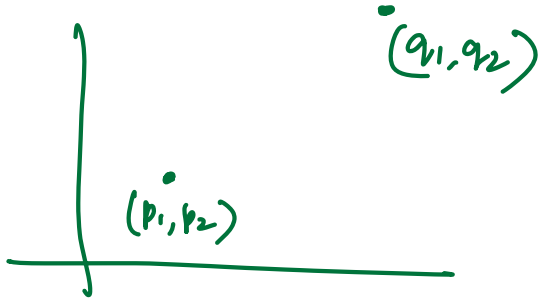
(5,5)

D	S	label	Distance	k=1	k=3
7	7	G	$\sqrt{(7-5)^2+(7-5)^2} = 2\sqrt{2}$		
6	4	B	$\sqrt{(6-5)^2+(4-5)^2} = \sqrt{2}$		B
7	4	G	$\sqrt{(7-5)^2+(4-5)^2} = \sqrt{5}$		
6	5	G	$\sqrt{(6-5)^2+(5-5)^2} = 1$	G	G
3	4	B	$\sqrt{(3-5)^2+(4-5)^2} = \sqrt{5}$		
1	4	B	$\sqrt{17}$		
4	3	B	$\sqrt{5}$		
3	5	B	2		B

Good | Bad

Distance:

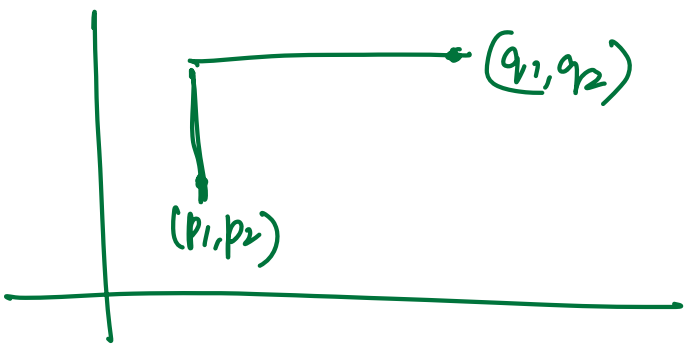
Euclidean Distance:



$$\sqrt{(q_1-p_1)^2 + (q_2-p_2)^2}$$

$$\left(\sum_{i=1}^{n} (q_i - p_i)^2 \right)^{1/2}$$

Manhattan Distance



$$(p_1 - q_1) + (p_2 - q_2)$$

$$\sum_{i=1}^{n} |p_i - q_i|$$

Minkowski Distance:

$$\left(\sum_{i=1}^{n} |p_i - q_i|^k \right)^{1/k}$$

$k = 2$

$k = 1$

Naive Bayes Classifier

Rwise:

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(A|B) \cdot P(B)$$

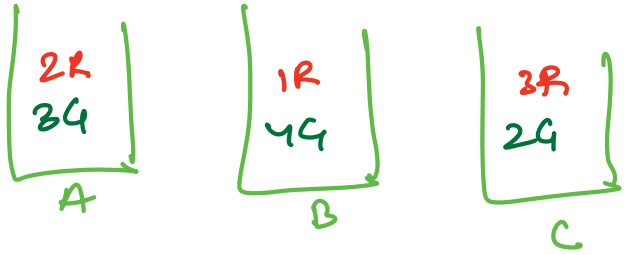
$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(B \cap A) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\boxed{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}}$$

Eg:



Q: Prob of getting a red ball given that bag A is chosen

$$P(R|A) = \frac{2}{5}$$

Q: Prob. of getting a red ball?

$$P(R) = P(R \cap A) + P(R \cap B) + P(R \cap C)$$

Q: Prob that bag A is chosen given that Red ball is drawn.

$$P(A|R) = \frac{P(A \cap R)}{P(R)}$$

$$= \frac{P(R \cap A)}{P(R)} = \frac{P(R|A) \cdot P(A)}{P(R)}$$

$$= \frac{P(R|A) \cdot P(A)}{P(R \cap A) + P(R \cap B) + P(R \cap C)}$$

(likelihood) conditional probability

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(A|B)$: posterior prob. $P(A)$: prior probability

Mushroom Dataset

Class: 1 2 3

features: shape, wt, color, radius ....

Test Mushroom?

$P(y=1|x) \rightarrow 0.25$

$P(y=2|x) \rightarrow 0.15$

$P(y=3|x) \rightarrow 0.6$

Sum: 1

max is 0.6

Test mushroom it belongs to class 3.

$$P(y=1) = \frac{n_1}{n_1+n_2+n_3}$$

$$P(y=2) = \frac{n_2}{n_1+n_2+n_3}$$

$$P(y=3) = \frac{n_3}{n_1+n_2+n_3}$$

$$P(y=1|x) = \frac{P(x|y=1) \cdot P(y=1)}{}$$

$P(x)$

$$P(y=1|x) = \frac{P(x|y=1) \cdot P(y=1)}{P(x \cap y=1) + P(x \cap y=2) + P(x \cap y=3)} \quad -$$

$$P(y=2|x) = \frac{P(x|y=2) \cdot P(y=2)}{P(x \cap y=1) + P(x \cap y=2) + P(x \cap y=3)} \quad -$$

$$P(y=3|x) = \frac{P(x|y=3) \cdot P(y=3)}{P(x \cap y=1) + P(x \cap y=2) + P(x \cap y=3)} \quad -$$

$$P(y=1|x) = \frac{P(x|y=1) \cdot P(y=1)}{Denom}$$

$$P(y=2|x) = \frac{P(x|y=2) \cdot P(y=2)}{Denom}$$

$$P(y=3|x) = \frac{P(x|y=2) \cdot P(y=2)}{Denom}$$

$P(y=1|x) \propto P(x|y=1) \cdot P(y=1)$ Prior Prob.

$P(y=2|x) \propto P(x|y=2) \cdot P(y=2)$

$P(y=3|x) \propto P(x|y=3) \cdot P(y=3)$

$P(x|y=1) = P(x_1|y=1) \cdot P(x_2|y=1) \cdot \ldots \cdot P(x_n|y=1)$

x is a test data point

$$P(x|y=1) = \prod_{i=1}^{n} P(x_i|y=1)$$

$$P(y=1|x) \propto \prod_{i=1}^{n} P(x_i|y=1) \cdot P(y=1)$$

$$P(y=c|x) \propto \prod_{i=1}^{n} P(x_i|y=c) \cdot P(y=c) \qquad c \in \{1,2,3\}$$

x_1 x_2 x_3 x_4 y

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$P(y=Yes) = \frac{9}{14}$$

$$P(y=No) = \frac{5}{14}$$

Outlook

	Yes	No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

$P(\text{outlook}=\text{Sunny} \mid y=\text{Yes})$

Humidity

	Yes	No
High	3/9	4/5
Normal	6/9	1/5

Temp

	Yes	No
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Windy

	Yes	No
Strong	3/9	3/5
Weak	6/9	2/5

Test Data Point

Outlook ! Sunny Temp ↓ Cool Humidity ↓ High Wind ! Strong

$P(y=Yes|x) = P(x|y=Yes) \cdot P(Yes)$

$= P(Outlook=Sunny|y=Yes) \cdot P(temp=Cool|y=Yes) \cdot$

$P(Humidity=High|y=Yes) \cdot P(wind=Strong|y=Yes) \cdot$

$P(y=Yes)$

$= \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14} = \frac{1}{9 \cdot 3 \cdot 7} = \frac{1}{189} = 0.0053$

$P(y=No|x) = P(x|y=No) \cdot P(No)$

$= P(Outlook=Sunny|y=No) \cdot P(temp=Cool|y=No) \cdot$

$P(Humidity=High|y=No) \cdot P(wind=Strong|y=No) \cdot$

$P(y=No)$

$= \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{14} = \frac{18}{675} = 0.0206$

Belongs to Class No.