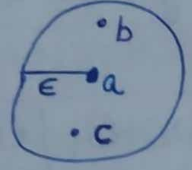
DBSCAN - Density Based Spatial Clustering of Application with Noise.

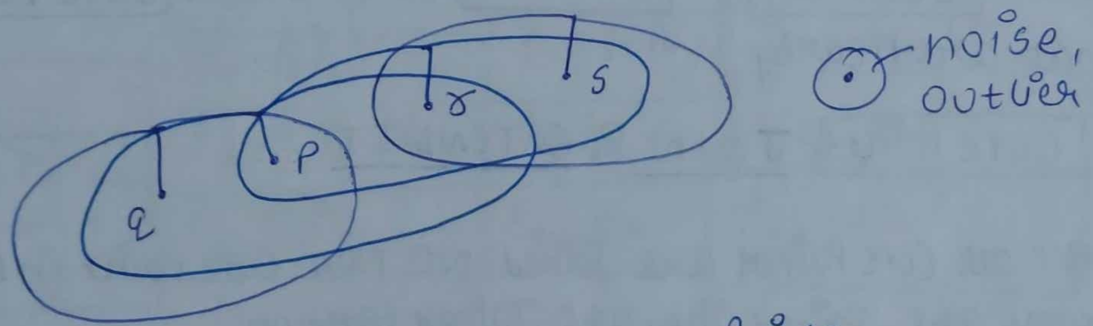
- Forms cluster on the Basis of $\boxed{\text{Density}}$
  - no. of Points in an defined area

- Parameters → ϵ = radius of circle | MinPoints = n | (3, 4, ... n)
  - To form a circle/area around a point



- Any Point which Contains Points ≥ MinPoints inside it's area (here we have Point 'a') = $\boxed{\text{CORE POINT}}$

eg



p, r = $\boxed{\text{CORE POINTS}}$
- center Points
- Contains Points ≥ Min Points

- If a Point is not a core Point But is a Neighbour of CORE POINT = BOUNDARY POINT

- If a Point is neither core nor Boundary = NOISE/OUTLIER

Directly Density Reachable →

From a to b | $\boxed{P \text{ to } q}$

(i) q must be an neighbor of P

(ii) P must be an CORE POINT

— X — X — X —

HOW CLUSTERS ARE FORMED

1. Take a random $\boxed{\text{CORE POINT}}$ and assign it to the first cluster

2. The $\boxed{\text{Core points}}$ closer to the first cluster / overlap its area or circle with $r = \epsilon$ are all added to the first cluster.

IMP

3. The $\boxed{\text{Core points}}$ that are CLOSER to the growing first cluster $\boxed{\text{JOIN IT}}$ & $\boxed{\text{EXTEND}}$ it to other $\boxed{\text{CORE POINTS}}$ that are nearby

$\boxed{\text{Core Points}}$ JOIN & EXTEND IT

$\boxed{\text{NOTE}}$ FIRST all Core Points are Joined, NO Non-Core Point in B/w the Process are Joined, They are Joined Later on

$\boxed{\text{NON CORE POINTS / Neighbor Points}}$ They are Joined in the first Cluster But the Cluster is not moved forward

$\boxed{\text{ONLY JOIN}}$

— X — X — X —

Numerical $\boxed{P1, P2, P3 \ldots P_{12}}$

$P_2 \rightarrow P_1$
$P_3 \rightarrow P_2, P_1$

P_4 — P_2, P_1, P_3

P_5 — P_2, P_1, P_3, P_4

$\rightarrow$

	P_1	P_2	. . . .	P_2
P_1				
P_2				
⋮				
P_{12}				

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
P_1	0											
P_2	1.41	0										
P_3	2.83	1.41	0									
P_4	4.24	2.83	1.41	0								
P_5	5.66	4.24	2.83	1.41	0							
P_6	5.83	4.47	3.16	2.00	1.41	0						
P_7	6.40	5.00	3.61	2.24	1.00	1.00	0					
P_8	5.83	4.47	3.16	2.00	1.41	2.83	2.24	0				
P_9	4.00	3.16	2.83	3.16	4.00	3.16	4.12	5.10	0			
P_{10}	1.41	2.00	3.16	4.47	5.83	5.66	6.40	6.32	3.16	0		
P_{11}	2.00	1.41	2.00	3.16	4.47	4.24	5.00	5.10	2.00	1.41	0	
P_{12}	3.16	2.83	3.16	4.00	5.10	4.47	5.39	6.00	1.41	2.00	1.41	0

$\boxed{\epsilon = 1.9}$

$\boxed{MinPts = 4}$

$x : \boxed{y, z}$ ← Distance < ϵ

$P_1 : P_2, P_{10}$	$P_4 : P_3, P_5,$	$P_7 : P_5, P_6$	$P_{10} : P_1, P_{11}$
$P_2 : P_1, P_3, P_{11}$	$P_5 : P_4, P_6, P_7, P_8$	$P_8 : P_5,$	$P_{11} : P_2, P_{10}, P_{12}$
$P_3 : P_2, P_4$	$P_6 : P_5, P_7$	$P_9 : P_{12}$	$P_{12} : P_9, P_{11}$

CORE POINTS → $\boxed{P_2, P_5, P_{11}}$

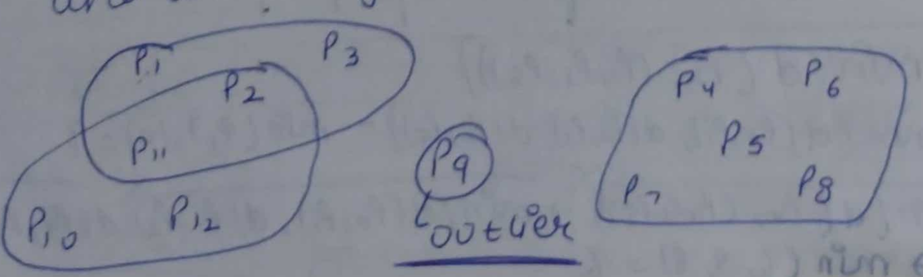
BOUNDARY POINTS → $P_1, P_3, P_4, P_6, P_7, P_8, P_{10}, P_{12}$

NOISE POINTS → $\boxed{P_9}$

IMP → If they were CP They would Extend cluster.

As P_4, P_6, P_7, P_8 are not core Points They will not extend the cluster and will only Joint | same with P_1, P_3, P_{10}, P_{12}.

Ans

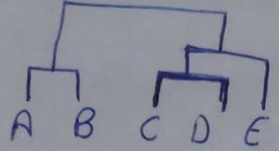


HIERARCHICAL CLUSTERING

→ Agglomerative (Bottom-up)

→ Divisive (TOP- DOWN)

→ Dendogram (Tree to represent Hierarchy of clusters)



Agglomerative Clustering

START = MINIMUM
In B/W = MINIMUM

SINGLE LINKAGE (minimum)

	P_1	P_2	P_3	P_4	P_5
P_1	0				
P_2	9	0			
P_3	3	7	0		
P_4	6	5	9	0	
P_5	11	10	2	8	0

- Always look for minimum Distance
  - combine them

	P_1	P_2	$[P_3, P_5]$	P_4
P_1	0			
P_2	9	0		
$[P_3, P_5]$	3	7	0	
P_4	6	5	8	0

Table 2
T_2

$\min(d(P_1, [P_3, P_5]))$
$\min(d(P_1,P_3), d(P_1,P_5))$
$\min(3, 11) = \underline{3}$ T_2

$\min(d(P_2, [P_3,P_5]))$,
$\min(d(P_2,P_3), d(P_2,P_5))$
$\min(7,10) = 7$ T_2

$\min(d(P_4, [P_3, P_5]))$ T_2
$\min(d(P_4,P_3), d(P_4,P_5))$
$\min(9, 8) = 8$

	$[P_1, P_3, P_5]$	P_2	P_4
$[P_1, P_3, P_5]$	0		
P_2	7	0	
P_4	6	5	0

Table 3
T_3

$\min(d(P_2, (P_1,P_3,P_5)))$ T_3
$\min(d(P_2,P_1), d(P_2,P_3), d(P_2,P_5)) = \min(9,7,10) = 7$

$\min(d(P_4, (P_1,P_3,P_5))) \rightarrow \min(d(P_4,P_1), d(P_4,P_3), d(P_4,P_5))$
$\rightarrow \min(6, 9, 8) = 6$ T_3

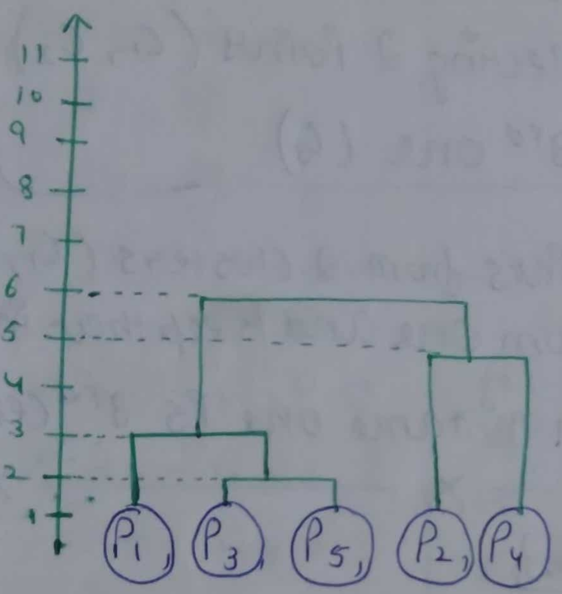
	[P_1,P_3,P_5]	[P_2,P_4]
[P_1,P_3,P_5]	0	
[P_2 P_4]	6	0

$\rightarrow$ (P_1, P_3, P_5, P_2, P_4)

$\min(d(P_2,P_4), d(P_1,P_3,P_5))$

$\min(d(P_2,P_1), d(P_2,P_3), d(P_2,P_5), d(P_4,P_1), d(P_4,P_3), d(P_4,P_5))$

$\min(9, 7, 10, 6, 9, 8) = 6$



COMPLETE LINKAGE $\boxed{\text{START = MINIMUM, INB/W = MAXIMUM}}$

Choose Maximum

Same matrix - START MINIMUM

	P_1	P_2	[P_3,P_5]	P_4
P_1	0			
P_2	9	0		
[P_3,P_5]			0	
P_4	6	5		0

$\rightarrow$ $\max((P_1,P_3),(P_1,P_5))$

$\max(3, 11) = 11$

SINGLE LINKAGE $\rightarrow$ START $\rightarrow$ MINIMUM
In B/W $\rightarrow$ MINIMUM

COMPLETE LINKAGE $\rightarrow$ START $\rightarrow$ MINIMUM
In B/W $\rightarrow$ MAXIMUM

STEPS IN K-Means ($K \geq 3$)

1. Randomly pick 1 centroid
2. Compute the Distance $D(x)$ of each Data Point (x) from the Cluster Center that was choosen.
3. Choose new Cluster Center from the Data Points with the probability of $\boxed{x \propto D(x)^2}$ (max Dist)

After Selecting 2 Points (C_1, C_2)
for 3rd one (C_3)

[Calc $D(x)$ of rem Data Points from 2 Clusters (C_1, C_2)
Only consider the minimum one and Keep that in List.]

[Out of that List the max Distance one is 3rd centroid C_3]

X ——— X ——— X

Kmeans

- calc distance of Data Points from C_1, C_2, C_3 and assign
- Data Point to that centroid where Distance is minimum.
- Recompute the centroid of cluster = mean of cluster $\left(\frac{x_1 + x_2}{n}, \frac{y_1 + y_2}{n}\right)$

X ——— X ———

Question

$K = 3$

A_1 (2,10)
A_2 (2,5)
A_3 (8,4)
A_4 (5,8)
A_5 (7,5)
A_6 (6,4)
A_7 (1,2)
A_8 (4,9)

* 1 Random Point = A_5 (7,5)

* Distance used → Manhattan Distance

* USING K-Means++ to select centroids for initial use

$A_1(2,10)$ | $A_2(2,5)$ | $A_3(8,4)$ | $A_4(5,8)$ | $A_5(7,5)$ | $A_6(6,4)$
$A_7(1,2)$ | $A_8(4,9)$

$A_5 - A_1 = 10$
$A_5 - A_2 = 5$
$A_5 - A_3 = 2$
$A_5 - A_4 = 5$
$A_5 - A_6 = 2$
$A_5 - A_7 = 9$
$A_5 - A_8 = 7$

$\boxed{C_2 = A_1 \text{ as } A_5 \text{ to } A_1 \rightarrow \max}$

D_1	D_2	min of D_1 and D_2
$A_5 - A_2 = 5$	$A_1 - A_2 = 5$	$A_5 - A_2 = 5$
$A_5 - A_3 = 2$	$A_1 \rightarrow A_3 = 12$	$A_5 - A_3 = 2$
$A_5 - A_4 = 5$	$A_1 \rightarrow A_4 = 5$	$A_5 - A_4 = 5$
	~~$A_1 \rightarrow A_5 = 10$~~	
$A_5 - A_6 = 2$	$A_1 \rightarrow A_6 = 10$	$A_5 - A_6 = 2$
$A_5 - A_7 = 9$	$A_1 \rightarrow A_7 = 9$	$A_5 - A_7 = 9$
$A_5 - A_8 = 7$	$A_1 \rightarrow A_8 = 3$	$A_1 - A_8 = 3$

$\boxed{C_3 = A_7}$

Three centroids for initial calculation

$C_1 \rightarrow A_5(7,5)$
$C_2 \rightarrow A_1(2,10)$
$C_3 \rightarrow A_7(1,2)$

A_5	A_1	A_7	min
$A_5 - A_2 = 5$	$A_1 \rightarrow A_2 = 5$	~~$A_7 \rightarrow A_1$~~ =	$A_7 - A_2$
$A_5 - A_3 = 2$	$A_1 \rightarrow A_3 = 12$	$A_7 \rightarrow A_2 = 4$	$A_5 - A_3$
$A_5 - A_4 = 5$	$A_1 \rightarrow A_4 = 5$	$A_7 \rightarrow A_3 = 9$	$A_1 - A_4$
$A_5 - A_6 = 2$	$A_1 \rightarrow A_6 = 10$	$A_7 \rightarrow A_4 = 10$	$A_5 - A_6$
[redacted] =	$A_1 \rightarrow A_8 = 3$	$A_7 \rightarrow A_6 = 7$	$A_1 \rightarrow A_8$
$A_5 - A_8 = 7$		$A_7 \rightarrow A_8 = 10$	

A_2 Belong to A_7

A_3, A_6 Belong to A_5

A_4, A_8 Belong to A_1

C_1: A_7 A_2

C_2: A_3 A_6 A_5

C_3: A_4 A_8 A_1

centroid of $C_1 \rightarrow (1.5, 3.5)$ x_1 | $C_2 \rightarrow (7, 4.3)$ x_2 | $C_3 \rightarrow (3.6, 9)$ x_3

x = Distance of all Points from x_1, x_2, x_3

x	from x_1	x from x_2	x from x_3	Belongs to
A_1	7	10.7	2.6	C_3
A_2	2	5.7	5.6	C_1
A_3	7	1.3	9.4	C_2
A_4	8	5.7	2.4	C_3
A_5	7	0.7	7.4	C_2
A_6	5	1.3	7.4	C_2
A_7	2	8.3	9.6	C_1
A_8	8	7.7	0.4	C_3

C_1: A_2 A_7 C_2: A_3 A_5 A_6 C_3: A_1, A_4, A_8 — **Ans**

No Changes in Cluster So No Calc → STOP

Classification Numerical

① Naive Bayes

1- $$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(y|x_1, x_2 \dots x_n) = \frac{P(x_1|y)\, P(x_2|y)\, P(x_3|y) \dots P(x_n|y) * P(y)}{P(x_1)\, P(x_2)\, P(x_3) \dots P(x_n)}$$

Numericals

$\underbrace{\text{Outlook} \mid \text{Temp} \mid \text{Humidity} \mid \text{windy}}_{x}$ $\underbrace{\boxed{\text{Class (Y/N)}}}_{y}$

STEP 1

- Count Probability of features of class Y
  - P(Yes) = 9/14
  - P(No) = 5/14

STEP 2

Calculate Probability of each unique feature of each ~~feo~~ col of X

- in outlook
  - Sunny → 5 times — Yes / No
  - Overcast → 4 times — Yes / No
  - Rain → 5 times — Yes / No

  } OUT of TOTAL Yes(9) and TOTAL No(5)

similarly for other features

Outlook	Y	N
Sunny	2/9	3/5
overcast	4/9	0
Rain	3/9	2/5

Temp	Y	N
hot	2/9	2/5
cool	3/9	1/5
mild	4/9	2/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5

WINDY	Y	N
False	6/9	2/5
True	3/9	3/5

WAITING FOR TEST DATA
after Building model

—— X —— X —— X ——

<rain, hot, high, false>

For Yes

$3/9 \times 2/9 \times 3/9 \times 6/9 \times 9/14 = 0.010582$

No

$2/5 \times 2/5 \times 4/5 \times 2/5 \times 5/14 = 0.018286$

∴ Ans is No

—— X —— X ——

Logistic Regression

* Converting 'y' from Linear into a value B/w 0 & 1 using sigmoid func^n. $\frac{1}{1+e^{-y}} \rightarrow \boxed{\text{sigmoid}(z) = \frac{1}{1+e^{-z}}}$ → S curve

$\boxed{y = mx+b} \rightarrow \boxed{y = \frac{1}{1+e^{-(mx+b)}}} \rightarrow \boxed{\log\left(\frac{y}{1-y}\right) = mx+b}$

"Logit" = "log odds"

$\boxed{P = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}} \rightarrow \log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$

$\text{Odds} = \frac{P(\text{event})}{1 - P(\text{event})}$ Instead of Y → we Take Probability P.

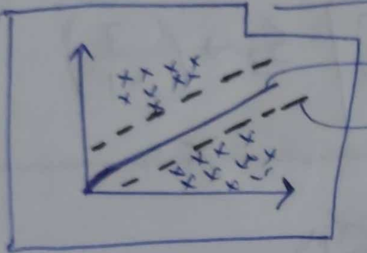
SVM

(1) Hyperplane to classify Data { +ve, -ve }

(2) 2 margines ‖ to Hyperplane

- 1st margin Passing through that +ve point which is closet to -ve Class
- 2nd margin Passing through that -ve point which is closet to +ve class

  - Closest to Opponent class



max marginal DIST

Marginal Distance → Distance B/w 2 margins → marginal DISTANCE

Support Vectors: nearest points towards respective Opponent class through which margins are drawn are known as Support vectors

---

SVM Kernels for Non linear seperable Data

- low D → High D

Eq of Hyperplane → (n-1)D | 2D-1D | 3D-2D

- $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n = 0$

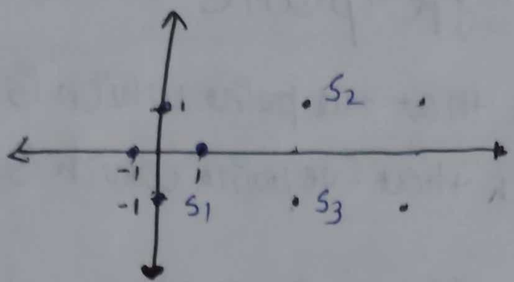
2D → line → $\boxed{\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0}$ → > 0 | < 0 | Plane

---

Linear → Hyperplane → $y = mx + b$, $y = wx + b$

+ve class → $(3,1)$ | $(3,-1)$ | $\begin{pmatrix}6\\1\end{pmatrix}$ | $\begin{pmatrix}6\\-1\end{pmatrix}$

-ve class → $(1,0)$ | $(0,1)$ | $\begin{pmatrix}0\\-1\end{pmatrix}$ | $\begin{pmatrix}-1\\0\end{pmatrix}$



Support vectors → $s_1 \rightarrow \begin{pmatrix}1\\0\end{pmatrix}$ | $s_2 \rightarrow \begin{pmatrix}3\\1\end{pmatrix}$ | $s_3 \rightarrow \begin{pmatrix}3\\-1\end{pmatrix}$

STEP 2 Add Bias (1) to find b later on

$s_1 \rightarrow \tilde{s}_1 \begin{pmatrix}1\\0\\1\end{pmatrix}$

$s_2 \rightarrow \tilde{s}_2 \begin{pmatrix}3\\1\\1\end{pmatrix}$

$s_3 \rightarrow \tilde{s}_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}$

STEP 3 FIND α

$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_1 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_1 = -1$ (as s_1 is on -ve side)
(opp class)

$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_2 = 1$

$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_3 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_3 + \alpha_3 \tilde{s}_3 \cdot \tilde{s}_3 = 1$

$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} = -1 \rightarrow \boxed{2\alpha_1 + 4\alpha_2 + 4\alpha_3 = -1}$

$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 1 \rightarrow \boxed{4\alpha_1 + 11\alpha_2 + 9\alpha_3 = 1}$

$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 1 \rightarrow \boxed{4\alpha_1 + 9\alpha_2 + 11\alpha_3 = 1}$

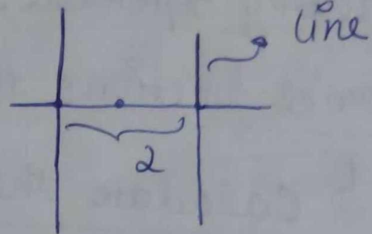
$\boxed{\alpha_1 = -3.5}$ $\boxed{\alpha_2 = 0.75}$ $\boxed{\alpha_3 = 0.75}$

[S3] Calculating weight vector ($\tilde{w}$) = $\sum \alpha_i \tilde{s}_i$ ($y = wx + b$)

↳ $-3.5\begin{pmatrix}1\\0\\1\end{pmatrix} + 0.75\begin{pmatrix}3\\1\\1\end{pmatrix} + 0.75\begin{pmatrix}3\\-1\\1\end{pmatrix}$

$= \boxed{\begin{pmatrix}1\\0\\-2\end{pmatrix}}$ as our vector was Biased

↳ $\boxed{w = \begin{pmatrix}1\\0\end{pmatrix}}$ $\boxed{b = -2}$ →



—— × —— × —— × ——

Non-Linear Separable

Soft margin → Try to find a line to seperate, but tolerate one or few misclassification.

1] The data is on $\boxed{\text{WRONG SIDE}}$ of the Decision Boundary but on the correct side or $\boxed{\text{on the margin}}$

2] The data is on the $\boxed{\text{WRONG SIDE}}$ of the Decision Boundary and on the $\boxed{\text{wrong side of margin}}$

- SVM tolerates a few data points to get misclassified and tries to balance the $\boxed{\text{TRADE-OFF}}$ b/w finding a line that maximizes the margin and minimizes the misclassification. 1

Degree of Tolerance → Penalty term given when SVM makes misclassification. $\boxed{\text{Higher the value}}$ - $\boxed{\text{Higher Penalty}}$ 1

SM → find a Hyperplane that maximize the margin while allowing some misclassification.

$\boxed{\text{degree of Tolerance}}$ (C)

Kernel Tricks →

- Transform the data into a higher-Dimensional space where it Becomes linearly seperable.
- Applying Kernel functions to the original feature space.
  - Calculate the DOT Product B/w 2 feats vectors in the Transformed Space without actually Computing the Transformation explicity.

SVM Kernels

Linear - linear seperable (original space) → $\boxed{X \cdot T * y}$ $\boxed{x^T y}$

Poly → non-linear | maps into Higher D using Poly func^n

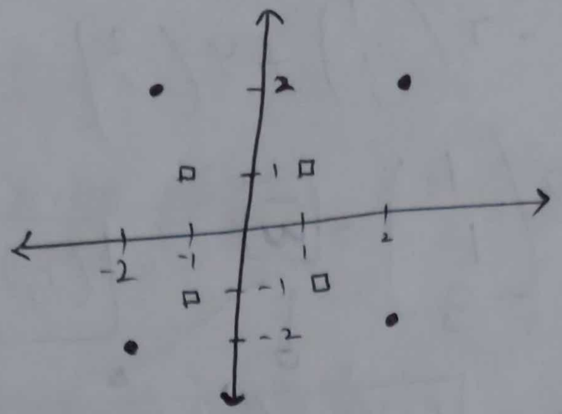
$(\gamma \cdot x^T y + r)^d$

rbf → maps into ∞ dimension space using Gaussian func^n → $\boxed{\exp(-\gamma \cdot \|x-y\|^2)}$

sigmoid → $\boxed{\tanh(\gamma \cdot X^T y + r)}$ using sigmoid

Numerical

+ve Class	-ve Class
$\boxed{\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}2\\-2\end{pmatrix}, \begin{pmatrix}-2\\-2\end{pmatrix}, \begin{pmatrix}-2\\2\end{pmatrix}}$	$\boxed{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}-1\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}}$

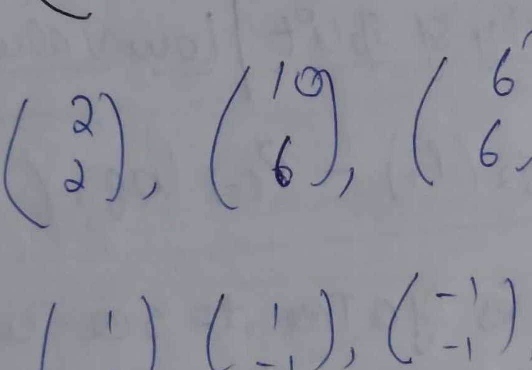


Conversion from lower D to Higher D

$$\phi_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

+ve class → $\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 10 \end{pmatrix}$

-ve class → $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}$



SV → $\boxed{\underset{s_1}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}} \text{ and } \underset{s_2}{\begin{pmatrix} 2 \\ 2 \end{pmatrix}}}$ → $s_1 \to \tilde{s}_1 \to \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$s_2 \to \tilde{s}_2 \to \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$

→ $\boxed{\begin{aligned} \alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_1 &= -1 \\ \alpha_1 \tilde{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 &= 1 \end{aligned}}$ → $\alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$

$\alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = 1$

→ $\boxed{\begin{aligned} 3\alpha_1 + 5\alpha_2 &= -1 \\ 5\alpha_1 + 9\alpha_2 &= 1 \end{aligned}}$ → $\boxed{\begin{aligned} \alpha_1 &= -7 \\ \alpha_2 &= 4 \end{aligned}}$.

$\boxed{\tilde{w} \Rightarrow}$ $\boxed{\sum_i \alpha_i \tilde{s}_i}$ $\Rightarrow -7\begin{pmatrix}1\\1\\1\end{pmatrix} + 4\begin{pmatrix}2\\2\\1\end{pmatrix}$

$\Rightarrow \begin{pmatrix}1\\1\\-3\end{pmatrix} = \boxed{\tilde{w}}$ $\boxed{(1,1) \text{ at } b=3}$

$\boxed{y \rightarrow wx+b} \rightarrow \boxed{w = \begin{pmatrix}1\\1\end{pmatrix} \text{ and } b = -3}$ *

$\boxed{\text{Decision Tree}}$ — Gain – Feature; Entropy → values of that feature

$\boxed{\text{Entropy} \rightarrow \text{Information Gain}}$

↳ measure of the purity of split | low value = Better

$\boxed{E \rightarrow H(S)} = \boxed{-P_{(+)}\log_2(P_+) - P_{(-)}\log_2(P_{(-)})}$

IG → avg of all entropies of a Tree to see which split is better. | High value = Better

$$Gain(\underbrace{S}_{\text{Subset}},\underbrace{A}_{\text{root}}) = H(S) - \sum_{v \in VAL} \frac{|S_v|}{|S|} H(S_v)$$

Numerical

① Entropy of whole dataset → $\boxed{\text{Outlook} \mid \text{Temp} \mid \text{Humidity} \mid \text{WIND}}$

↳ $\boxed{\text{Y feature}}$ → Yes – 9, → NO – 5] 14

$$E(S) = -\frac{9}{14}\log_2\left(+\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.94$$

---

② Entropy of values of each feature

Outlook

Sunny → 2 Yes, 3 No → $E(S_{sunny}) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$

$= \underline{0.971}$

Overcast → 4 Yes, 0 No → $E(S_{overcast}) = -\frac{4}{4}\log_2\left(\frac{4}{4}\right) = \underline{0}$

Rain → 3 Yes, 2 No → $E(S_{Rain}) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$

$= \underline{0.971}$

$$Gain(S, outlook) = Entropy(S) - \sum_{V \in (Sunny, O, R)} \frac{|S_V|}{|S|} Entropy(S_V)$$

$$\boxed{Entropy(S) - \frac{5}{14} Entropy(S_{sunny}) - \frac{4}{14} Entropy(S_O) - \frac{5}{14} Entropy(S_R)}$$

$= \boxed{0.2464}$

Temp

Hot → 2Y, 2N → $-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = \underline{1}$

mild → 4Y, 2N → $-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) = \underline{0.9183}$

Cool → 3Y, 1N → $-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = \underline{0.8113}$

$$Gain(S, Temp) = Entropy(S) - \frac{4}{14} Entropy(S_{Hot}) - \frac{6}{14} Entropy(S_m) - \frac{4}{14} Entropy(S_c)$$

$= \underline{0.0289}$

Humidity

→ High $\begin{cases} 3Y \\ 4N \end{cases}$ → $E(S_H) = -\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right) = 0.9852$

Normal $\begin{cases} 6Y \\ 1N \end{cases}$ → $E(S_N) = -\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right) = 0.5916$

$Gain(S, H) \Rightarrow E(S) - \frac{7}{14}E(S_H) - \frac{7}{14}E(S_N)$

$\Rightarrow$ 0.1516

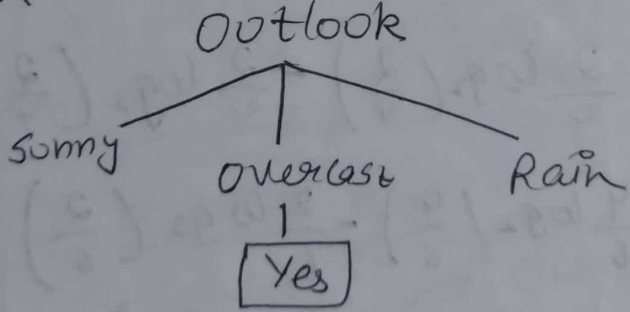
WIND

weak $\begin{cases} 6Y \\ 2N \end{cases}$ → $-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 0.8113$

Strong $\begin{cases} 3Y \\ 3N \end{cases}$ → $-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{4}{6}\right) = 1$

$Gain(S, w) = E(S) - \frac{8}{14}E(S_w) - \frac{6}{14}E(S_S) = 0.048$

OUTLOOK has highest IG



2nd split

- Take only those records where Sunny is present and repeat the STEPS
- SIMILAR for RAIN