ML

2017-2018 →

1.) Genetic algo
2.) NN → Back prop & W update
3.) SVM in case of non-linear data
   - overfitting
   - Primal & dual problems
4.) KNN
5.) LR, DT, NN, KNN → classification.
6.) Locally weighted Regression.

2016-2017

1.) NN - 2 perceptron problem
2.) Hypothesis, plot, Genetic algo
3.) Decision tree
4.) Prob, SVM,
5.) Gradient descent.
   - learning rate

- Variance, Training Set, Bias
- K-means, outliers - robustness
- Perceptron training
- CROSSOVER - GA
- KNN numerical
- DT

- Belief NW -

$$-1 \times \log(0.97) - 0\times = 0.029$$

$$-0\times \quad - (1)\log(0.001)$$

SVD

Given $A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}$ $A^T = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}$ $A^TA = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}$

$$\det(A^TA - \lambda I) = \begin{vmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{vmatrix}$$

$$= \lambda^2 - 100\lambda + 1600$$

$$= (\lambda - 20)(\lambda - 80)$$

Considering $\lambda = 20$.

$(A^TA - \lambda I) = \cancel{0}$

$$\begin{bmatrix} 6 & 18 \\ 16 & 54 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0$$

$6y_1 + 18y_2 = 0 \Rightarrow y_1 + 3y_2 = 0$

$16y_1 + 54y_2 = 0 \Rightarrow y_1 + 3y_2 = 0$

$y_1^2 + y_2^2 = 1$

$10y_2^2 = 1$

$y_2 = \frac{1}{\sqrt{10}}$ $y_1 = \frac{-3}{\sqrt{10}}$

Considering $\lambda = 80$

$$\begin{bmatrix} -54 & 18 \\ 18 & -6 \end{bmatrix} \times \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -3 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0$$

$-3z_1 + z_2 = 0$ $z_2 = 3z_1$

$3z_1 - z_2 = 0$

$z_1^2 + z_2^2 = 0$ $z_1 = \frac{1}{\sqrt{10}}$, $z_2 = \frac{3}{\sqrt{10}}$

$AV = U\Sigma$

$$\begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} = U \times \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}$$

1.061 1.577
0.516
0.357
0.286

$$\begin{bmatrix} 2\sqrt{10} \\ 2\sqrt{10} \end{bmatrix} \begin{bmatrix} -\sqrt{10} \\ \sqrt{10} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}$$

Bayes optimal classifier.

Given.

$P(h_1|D) = 0.4$

$P(h_2|D) = 0.3$

$V = \{\oplus, \ominus\}$ $\quad P(h_3|D) = 0.3$

$P(h_i/D)$ \ class	$\ominus$	$\oplus$
0.4	0	1
0.3	1	0
0.3	1	0
	0.6	0.4

$\Rightarrow$ Classified as $\ominus$ class.

$P(h_1/D) = 0.4$

$P(h_2|D) = 0.3$

$P(h_3|D) = 0.3$

$P(h_4|D) = 0.6$

$P(h_5|D) = 0.7$

$P(h_6|D) = 0.5$

$P(h_7|D) = 0.3$

$h_1(x) = B$

$h_2(x) = C$

$h_3(x) = B$

$h_4(x) = T$

$h_5(x) = B$

$h_6(x) = C$

$h_7(x) = T$

	B	C	T
h_1	✓		
h_2		✓	
h_3	✓		
h_4			✓
h_5	✓		
h_6		✓	
h_7			✓
	1.4	0.8	0.9

Perceptron Training Rule

x_1	x_2	Target	output
0	0	1	$-0.2 \approx 0$
0	1	1	$0.4 \approx 1$
1	0	0	$-0.5 \approx 0$
1	1	1	$0.1 \approx 1$

$X_0 = -1$

$w_0 = 0.2$

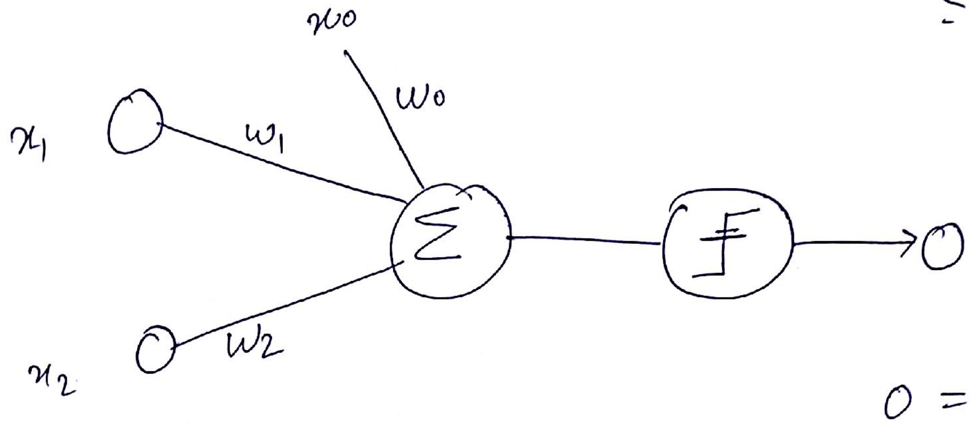
$w_1 = -0.3$

$w_2 = 0.6$

$b = 0.1$

$\Sigma x_i * w_i > 0 \quad = 1$

$\leq 0 \quad = 0$



$$O = \begin{cases} 1 & , \Sigma w_i x_i > 0 \\ -1 & , \text{other} \end{cases}$$

$$w_0 x_0 + w_1 x_1 + w_2 x_2 = \text{target}$$

$0\ 0 \rightarrow -0.2$

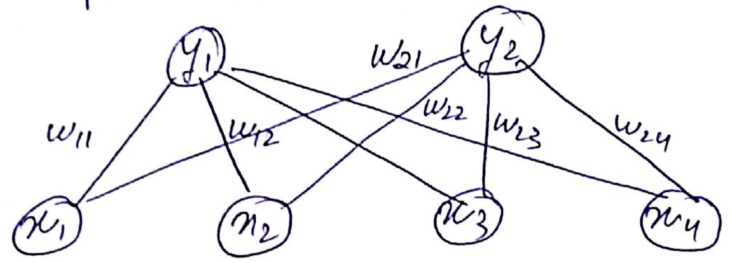
$0\ 1 \rightarrow -0.2 + 0.6 = 0.4$

$1\ 0 \rightarrow -0.2 - 0.3 = -0.5$

$1\ 1 \rightarrow -0.2 - 0.3 + 0.6 = 0.1$

Self-Organising Map.

- 2 clusters.
- 4 Input vectors



I/P vectors = 4

O/P vector = 2.

STEP 1 - Initialize weights b/w 0 & 1

$$w_{ij} = \begin{array}{cc} Y_1 & Y_2 \end{array}$$
$$w_{ij} = \begin{bmatrix} 0.2 & 0.9 \\ 0.4 & 0.7 \\ 0.6 & 0.5 \\ 0.8 & 0.3 \end{bmatrix}$$

STEP 2 → First input vector $X = [0 \; 0 \; 1 \; 1]$

Calculate euclidean distance.

$$[X - Y_1] \; , \; D_{x_1 - Y_1} = \Sigma(w_{ij} - x_i)^2 = (0.2-0)^2 + (0.4-0)^2 + (0.6-1)^2 + (0.8-1)^2$$
$$= 0.4 \Rightarrow \underline{\text{WINNER}}$$

$$D_{x_1 - Y_2} = (0.9-0)^2 + (0.7-0)^2 + (0.5-1)^2 + (0.3-1)^2 = 2.04$$

STEP-3 → update weight.

$$w_{ij}(new) = w_{ij}(old) + \alpha[x_i - w_{ij} old]$$

$$w_{11} = 0.2 + 0.5[0 - 0.2] = 0.1$$
$$w_{12} = 0.4 + 0.5[0 - 0.4] = 0.2$$
$$w_{13} = 0.6 + 0.5[1 - 0.6] = 0.8$$
$$w_{14} = 0.8 + 0.5[1 - 0.8] = 0.9$$

Scanned by CamScanner

So, weights changes for y_1.

Now, weight

$$w_{ij} = \begin{array}{cc} X_1 & X_2 \\ \left[\begin{matrix} 0.1 & 0.9 \\ 0.2 & 0.7 \\ 0.8 & 0.5 \\ 0.9 & 0.3 \end{matrix}\right. & \left.\right] \end{array}$$

Second input vector [1 0 0 0]

$$D_{x2 \to y1} = \Sigma (w_j - x_{ij})^2 = (0.1-1)^2 + (0.2)^2 + (0.8)^2 + (0.9)^2$$
$$= 2.3$$

$$D_{x2 \to y2} = \Sigma (w_j - x_{ij})^2 = (0.9-1)^2 + (0.7)^2 + (0.5)^2 + (0.3)^2$$
$$= 0.84 \quad \underline{\underline{\text{WINNER}}}$$

Update weights of y_2 $\begin{bmatrix} 0.9 \\ 0.7 \\ 0.5 \\ 0.3 \end{bmatrix}$

$$w_{21} = w_{21\,old} + \alpha [x_i - w_{21\,old}]$$

$$w_{21} = 0.9 + 0.5 [1 - 0.9] = 0.95$$

$$w_{22} = 0.7 + 0.5 [-0.7] = 0.35$$

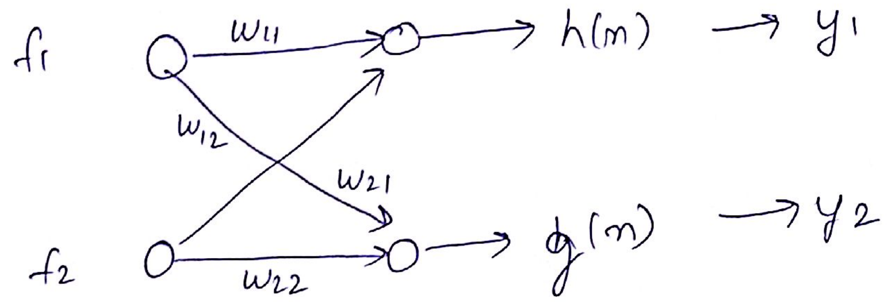
$$w_{23} = 0.5 + 0.5 [-0.5] = 0.25$$

$$w_{24} = 0.3 + 0.5 [-0.3] = 0.15$$

Update weights.

$$w_{ij} = \begin{bmatrix} 0.1 & 0.95 \\ 0.2 & 0.35 \\ 0.8 & 0.25 \\ 0.9 & 0.15 \end{bmatrix}$$

① → 2 perceptron model



(a) $y_1 = h(f_1 \times w_{11} + f_2 \times w_{21})$
$y_2 = g(f_1 \times w_{12} + f_2 \times w_{22})$

f_1	f_2	O/P	
1	1	0	00
1	4	1	01
2	5	1	01
1	2	0	00
4	2	2	10
4	5	3	11
5	4	3	11
5	2	2	10

$h(w_{11} + w_{21}) = 0$, $g(w_{12} + w_{22}) = 0$

$\Rightarrow w_{11} + w_{21} < 0$
$w_{12} + w_{22} < 0$

$h(w_{11} + 4w_{21}) = 0$, $g(w_{12} + 4w_{22}) = 1$

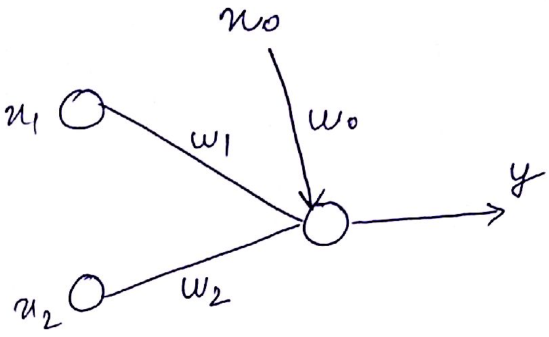
$w_{11} + 4w_{21} < 0$
$w_{12} + 4w_{22} > 0$

$\Rightarrow w_{11} = -5 \quad w_{21} = +1$
$w_{12} = -3 \quad w_{22} = 1$

Scanned by CamScanner

Design a NN for.

x_1	x_2	class
0	0	0
0	1	0
1	0	1
1	1	0



Equation → $y = 1 \quad w_0 + \cancel{w_1 x_1 + w_2 x_2 > 0}$
$= -1 \qquad\qquad \leq 0$

$w_0 + (0)\cancel{w_1} + w_2(0) < 0$

$$y = \begin{cases} 1 & \text{if } \sum w_i x_i \geq 0 \\ 0 & \sum w_i x_i < 0 \end{cases}$$

With $x_1 \; x_2$

$0 \; 0 \rightarrow 0 \Rightarrow \quad w_0 + w_1(0) + w_2(0) < 0 \quad \Rightarrow w_0 = -1$

$0 \; 1 \rightarrow 0 \Rightarrow \quad w_0 + w_2 < 0 \quad \Rightarrow w_2 = -1$

$1 \; 0 \rightarrow 1 \Rightarrow \quad w_0 + w_1 \geq 0 \quad \Rightarrow w_1 = 1.5$

$1 \; 1 \rightarrow 0 \Rightarrow \quad w_0 + w_1 + w_2$
$-1 \quad -1 \quad +1.5 \quad = < 0 \Rightarrow$ valid,

So, $(w_0, w_1, w_2) = [-1 \;\; -1 \;\; 1.5]$

Q. Learning

a 0	b 0 →	c 100
d 0	e 0	f ↑

$\gamma = 0.9$

(1)

	←	↑	↓	→
a	0	0	0	0
b	0	0	0	$\boxed{100}$
c	\	\	\	\
d	0	0	0	0
e	0	0	0	0
f	0	$\boxed{100}$	0	0

(2) Update value of B → $0 + 0.9(100) = 90$

f → $0 + 0.9(100) = 90$

a 0	b 90	c 100
d 0	e 0	f 90

	←	↑	↓	→
a	0	0	0	90
b	0	0	0	100
c	\	\	\	\
d	0	0	0	0
e	0	90	0	90
f	0	100	0	0

(3) Update value of a → $0 + 0.9 \times 90 = 81$

" e → $0 + 0.9 \times 90 = 81$

a 81	b 90	c 100
d 0	e 81	f 90

	←	↑	↓	→
a	0	0	0	90
b	81	0	81	100
c	/	/	/	/
d	0	81	0	81
e	0	90	0	90
f	81	100	0	0

(4) Update value of d

→ $0 + 0.9 \max(81, 81)$ 72.9

a 81	b 90	c 100
d 73	e 81	f 90

FINDING SVD

$$A = U \Sigma V^T$$

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}$$

Step 1 → Find $A^TA = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}\begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}$

Step 2 → Find $\det|A^TA - \lambda I| = \begin{vmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{vmatrix}$

$$= \lambda^2 - 100\lambda + 1600$$

Equating to zero, $\lambda^2 - 100\lambda + 1600 = 0$

$$\lambda_1 = 20, \quad \lambda_2 = 80$$

Putting $\lambda = 80$ in $A^TT - \lambda I = 0$

$$\begin{bmatrix} -54 & 18 \\ 18 & -6 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow -54y_1 + 18y_2 = 0 \text{ \& } 18y_1 - 6y_2 = 0.$$

$$\Rightarrow y_1 = 1, y = 3$$

Normalized values $y_1 = \frac{1}{\sqrt{10}}$, $y_2 = \frac{3}{\sqrt{10}}$ $\Rightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix}$

Similarly using $\lambda_2 = 20$, we get $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$

$$V = \begin{bmatrix} 1/\sqrt{10} & -\frac{3}{\sqrt{10}} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{bmatrix} = \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}$$

Using formula, $AV = U\Sigma$

$$\begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{bmatrix} = U \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}$$

$$A^{-1} = \frac{Adj(A)}{|A|} \Rightarrow \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix} \quad Cof = \begin{bmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{bmatrix}$$

$$|A| = 4\sqrt{5} \times 2\sqrt{5} = 40$$

$$A^{-1} = \begin{bmatrix} \frac{2\sqrt{5}}{40} & 0 \\ 0 & \frac{4\sqrt{5}}{40} \end{bmatrix} = [$$

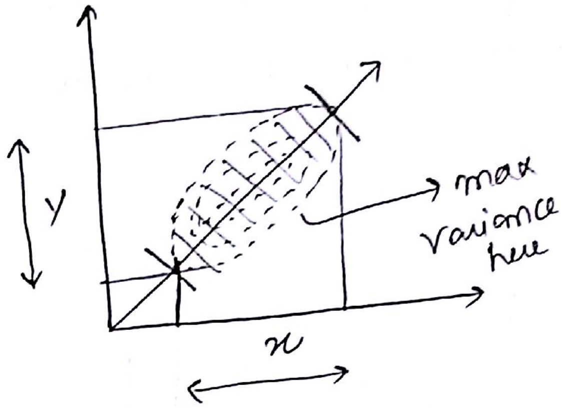
$$U = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{bmatrix}\begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}^{-1}$$

$$U = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

PRINCIPAL COMPONENT ANALYSIS (PCA) → EIGEN PROBLEM.

$det(Cov x - \lambda I) = 0$

- Find direction of maximum variance.
- Find orthogonal directions



Question →

Let $A^T = \begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix} \begin{matrix} \to x \\ \to y. \end{matrix}$

$\bar{x} = 0.5$

$\bar{y} = 2.125$

$$Cov = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

X	Y	A $x_i - \bar{x}$	B $y_i - \bar{y}$	AB	A^2	B^2
2	4	1.5	1.875	2.8125	2.25	3.556
1	3	0.5	0.875	0.4375	0.25	0.7656
0	1	-0.5	-1.125	0.5625	0.25	1.2656
-1	0.5	-1.5	-1.625	0.4375	2.25	2.6406
				6.25	5	8.1874

$Cov(x, x) = \frac{5}{3} = 1.67$

$Cov(y, y) = \frac{8.1874}{3} = 2.73$

$Cov(x, y) = Cov(y - x) = \frac{6.25}{3} = 2.083$

$$S = Cov = \begin{array}{c|cc} & X & Y \\ \hline X & 1.67 & 2.083 \\ Y & 2.083 & 2.73 \end{array}$$

$$|S - \lambda I| = 0$$

$$\left| \begin{bmatrix} 1.67 & 2.083 \\ 2.083 & 2.73 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\begin{vmatrix} 1.67 - \lambda & 2.083 \\ 2.083 & 2.73 - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - 4.4\lambda + 0.2202 = 0$$

$$\boxed{\begin{aligned} \lambda_1 &= 4.3494 \\ \lambda_2 &= 0.0506 \end{aligned}}$$

Consider λ_1,

$$\begin{bmatrix} 1.67 - 4.3494 & 2.083 \\ 2.083 & 2.73 - 4.349 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} = 0$$

$$\boxed{a_{11}^2 + a_{12}^2 = 1}$$ using this.

$a_{11} = 0.61$
$a_{12} = 0.79$

Consider λ_2.

$a_{21} = 0.79$
$a_{22} = 0.61$

Principle components: $\boxed{Z_1 = a_{11} x_1 + a_{12} x_2}$

$\boxed{Z_2 = a_{21} x_1 + a_{22} x_2}$

	Size	Color	Shape	Level
→ 1	Big	Red	Circle	−ve
→ 2	Small	Red	Triangle	−ve
3	Small	Red	Circle	+ve
→ 4	Big	Blue	Circle	−ve
5	Small	Blue	Circle	+ve

→ G_0 → Specilize

S_0 → Generalise

G_i → Conflict − remove

$\boxed{S_0}$ < ϕ ϕ ϕ >

x_1 −ve S_1 < ϕ ϕ ϕ >

x_2. −ve S_2 < ϕ ϕ ϕ >

x_3 +ve S_3 < Small Red Circle >

x_4 −ve S_4 < Small Red Circle >

x_5 +ve S_5 < Small ? Circle >

→ Answer −

x_5 +ve G_5 < Small ? Circle >

x_4 −ve G_4 < Small ? Circle >

x_3 − +ve G_3 < Small ? Circle >

x_2 −ve G_2 < Small ? Circle > ~~< ? Blue ? >~~ ~~< Big ? Triangle >~~

x_1 −ve G_1 < Small, ? ? > < Blue, ? ? > < ? ? Triangle >

x $\boxed{G_0}$ < ? ? ? >

MDL

Given 10 values, 4 subset

$$L(E) = \log({}^{10}C_4 \times 2) = 8.715 \times 14 = 122.01$$

Given

H_1

A \ P	Y	N
Y	7	2
N	0	5

H_2

A \ P	Y	N
Y	8	1
N	0	5

$$L(E|H_1) = \underbrace{\log({}^{7}C_0)}_{FN} + \underbrace{\log({}^{7}C_2)}_{FP} = 4.39$$

$$L(E|H_2) = \log({}^{8}C_0) + \log({}^{6}C_1) = 2.59$$

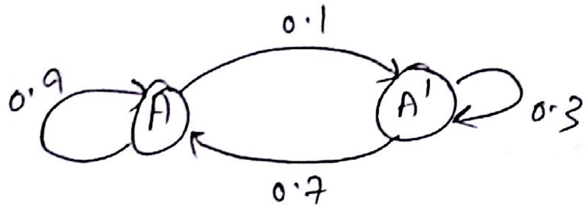
$$\text{Compression } H_1 = L(E) - L(H_1) - L(E|H_1) = 108.81$$

$$\text{Compression } H_2 = L(E) - L(H_2) - L(E|H_2) = 105.42$$

H_1 is better.

Given Initial State.

$$S_0 = [0.2 \quad 0.8] \qquad A = \begin{bmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{bmatrix}$$



Given

A =

	H	C
H	0.7	0.3
C	0.4	0.6

B =

	S	M	L
H	0.1	0.4	0.5
C	0.7	0.2	0.1

$$\pi = \begin{array}{cc} H & C \\ [0.6 & 0.4] \end{array}$$

Observation → S M S L

$$P(HHCC) = P(H_0) \times P(S|H) \times P(H|H) \times P(M|H) \times P(C|H) \times P(S|C) \times P(C|C) \times P(L|C)$$

$$= 0.6 \times 0.1 \times 0.7 \times 0.4 \times 0.3 \times 0.7 \times 0.6 \times 0.1$$

$$= \underline{0.000212}$$

Scanned by CamScanner

$$\theta_A^{k}(1-\theta_A)^{n-k}$$

Given

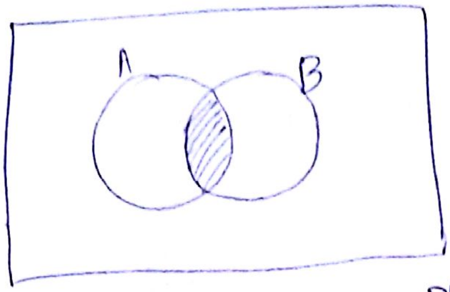
	$\theta_B = 0.5$ $\theta_A = 0.6$		$P(k_A)$	$P(k_B)$	Normalized	
1	5H	5T	0.000976	0.00079	0.44	0.55
2	9H	1T	0.00403	0.000976	0.80	0.20
3	8H	2T	0.00269	0.000976	0.73	0.27
4	4H	6T	0.000530	0.000976	0.35	0.65
5	7H	3T	0.00179	0.000976	0.65	0.35

Now,

Coin A	Coin B
$0.45 \times (5,5) = 2.2, 2.2$	
$0.80 \times (9,1) = 7.2, 0.8$	
$0.73 \times (8,2) = 5.9, 1.8$	
$0.35 \times (4,6) = 1.4, 2.1$	
$0.65 \times (7,3) = 4.5 \quad 1.9$	
$21.3, 8.6$	$11.7 \quad 8.4$

$$\hat{\theta}_A = \frac{21.3}{21.3+8.6} = 0.71$$

$$\hat{\theta}_B = 0.58$$



$P(A,B)$
$= P(A) \times P(B|A)$
$= P(B) \times P(A|B)$

Why 100? why not 10000?

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

Bayes Theorem

Example: Three companies A, B and C makes 35%, 35% and 30% of all the lamps in market. Probability of their lamp being defective is 1.5%, 1% and 2% respectively. What is the probability that a randomly selected defective lamp was manufactured in factory C?

Bayes Theorem

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

Posterior, Prior, Likelihood and Evidence ($P(A|B)$, $P(A)$, $P(B|A)$, $P(B)$)

What we want is $P(C|D)$ it is $\frac{P(D|C)P(C)}{P(D)}$

$P(D) = P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) =$ → Probability of defective

$.35 * 0.015 + .35 * 0.01 + .3 * 0.02 = 0.01475$

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{0.02 * 0.3}{0.01475} = 0.407$$

Hypothesis

X	Y
10	0
11	0
12	0
13	1
14	0
15	1
16	0
17	1
18	1

h_1	h_2	...
0	1	...
0	0	...
0	1	...
1	0	...
1	1	...
1	0	...
1	1	...
1	0	...
1	1	...

In this example h_1, h_2, ... are hypothesis.
Hypothesis is a function that aims to provide value of the Y
Can you identify h_1 and h_2
Represent H as candidate set of hypothesis, *i.e.* $h_i \in H$
Size of H is at least 2^m

Example- Solution

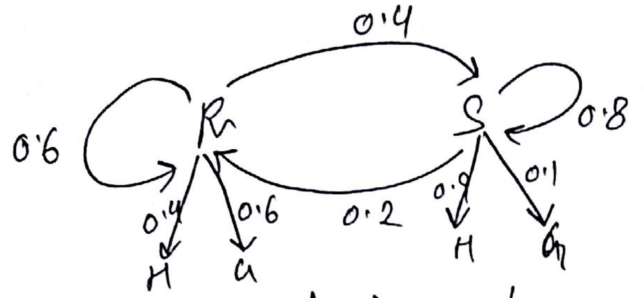
Given $P(A) = 0.35$, $P(B) = 0.35$, $P(C) = 0.30$
$P(D|A) = 0.015$, $P(D|B) = 0.01$, $P(D|C) = 0.02$.

We need to calculate $P(C|D) = \frac{P(C)\ P(D|C)}{P(D)}$

Now, $P(D) = P(D|A) * P(A) + P(D|B) * P(B) + P(D|C) * P(C)$
$= 0.01475$

$$P(C|D) = \frac{0.02 * 0.3}{0.01475} = 0.407.$$

Hidden Markov model + Bayes Rule.



$A =$

	R	S
R	0.6	0.4
S	0.2	0.8

Let $P(R_0) = \frac{1}{2}$

$P(S_0) = \frac{1}{2}$

$B =$

	H	G
R	0.4	0.6
S	0.9	0.1

Suppose I am happy H_1, what is probability ~~that~~ $P(R_1|H_1)$

$$P(R_1|H_1) = \frac{P(H_1|R_1) \times P(R_1)}{P(H_1)}$$

$$P(R_1) = P(R_1|R_0) \cdot P(R_0) + P(R_1|S_0) \cdot P(S_0)$$

$$= 0.6 \times 0.5 + 0.2 \times 0.5 = 0.4$$

$$P(S_1) = P(S_1|S_0) \times P(S_0) + P(S_1|R_0) \times P(R_0) = 0.8 \times 0.5 + 0.4 \times 0.5 = 0.6$$

$$P(H_1) = P(H_1|R_1) \cdot P(R_1) + P(H_1|S_1) \times P(S_1)$$

$$= 0.4 \times 0.4 + 0.9 \times 0.6.$$

$$= 0.16 + 0.54 = 0.7$$

$$P(R_1|H_1) = \frac{0.4 \times 0.4}{0.7} = 0.2286$$

Scanned by CamScanner

For example on last Page :

$$B = \begin{array}{c} \\ H \\ C \end{array}\begin{array}{c} \begin{array}{ccc} S & M & L \end{array} \\ \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array} \qquad A = \begin{array}{c} \\ H \\ C \end{array}\begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

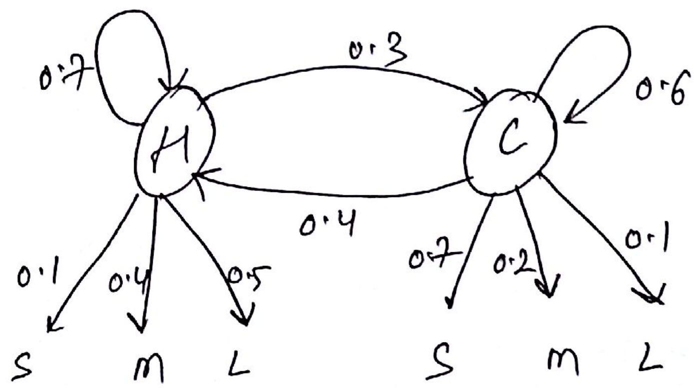
$$\pi = \begin{array}{c} \begin{array}{cc} H_0 & C_0 \end{array} \\ \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \end{array}$$

let observation on S, M, S, L.

$$P(H\,H\,C\,C) = P(S|H) \times P(H_0) \times P(M|H) \times P(H_1|H_0) \times P(S|C) \times P(C_1|H_1) \times P(L|C) \times P(C|C)$$

$$= (0.1 \times 0.6) \times (0.4 \times 0.7) \times (0.7 \times 0.3) \times (0.1 \times 0.6)$$

$$= 0.000212$$



2 types of Question

① Say Coverage was S, calculate $P(H_1/S_1)$

② Given observation S M S L, calculate $P(HHCC)$

Day	Outlook	Temp	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	N
D2	S	H	H	Strong	N
D3	Overcast	H	H	W	Y
D4	Rainy	Mild	H	W	Y
D5	R	Cool	Normal	W	Y
D6	R	C	N	S	N
D7	O	C	N	S	Y
D8	S	M	H	W	N
D9	S	C	N	W	Y
D10	R	M	N	W	Y
D11	S	M	N	S	Y
D12	O	M	H	S	Y
D13	O	H	N	W	Y
D14	R	M	H	S	N

Entropy [9+, 5-]

$$= -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= 0.94$$

INFORMATION GAIN

Information Gain of an attribute is the expected reduction in entropy caused by positioning the examples according to that attribute

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

here S_v contains that data items of S where value of attribute A is v

$S[9+, 5-]$, $E = 0.940$

Humidity

High → $S[3+, 4-]$, $E = 0.985$

Normal → $S[6+, 1-]$, $E = 0.592$

$S[9+, 5-]$, $E = 0.940$

Wind

Weak → $S[6+, 2-]$, $E = 0.811$

Strong → $S[3+, 3-]$, $E = 1.000$

$$Gain(S, Humidity) = 0.940 - \left(\frac{7}{14}\right)0.985 - \left(\frac{7}{14}\right)0.592 = 0.151$$

$$Gain(S, Wind) = 0.940 - \left(\frac{8}{14}\right)0.811 - \left(\frac{6}{14}\right)1.000 = 0.048$$

Entropy(S_v)

INFORMATION GAIN and DECISION TREE

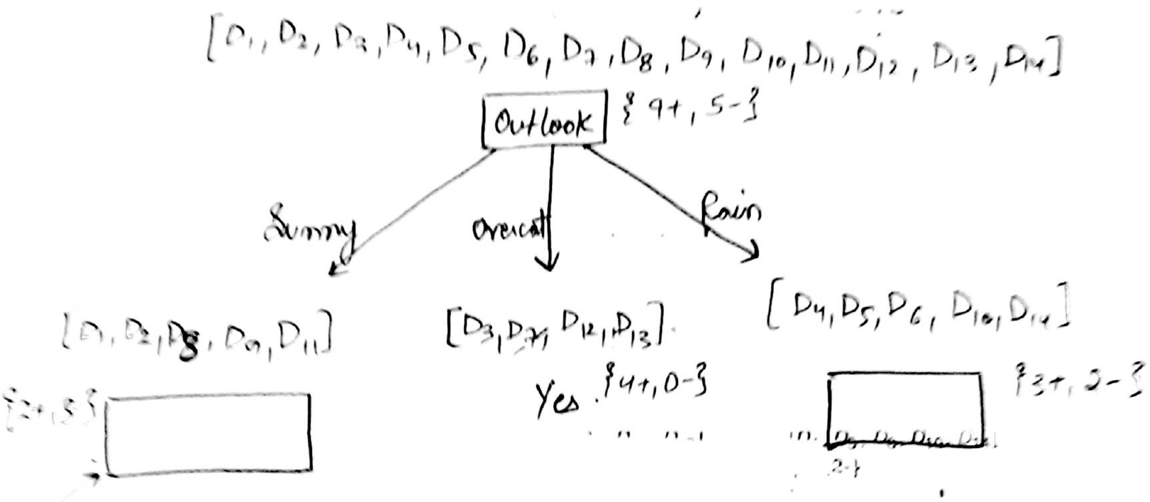
Gain(S, Humidity) = 0.151

Gain(S, Wind) = 0.048

Gain(S, Outlook) = 0.246 →

Gain(S, Temperature) = 0.029

$$Gain(S,A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$



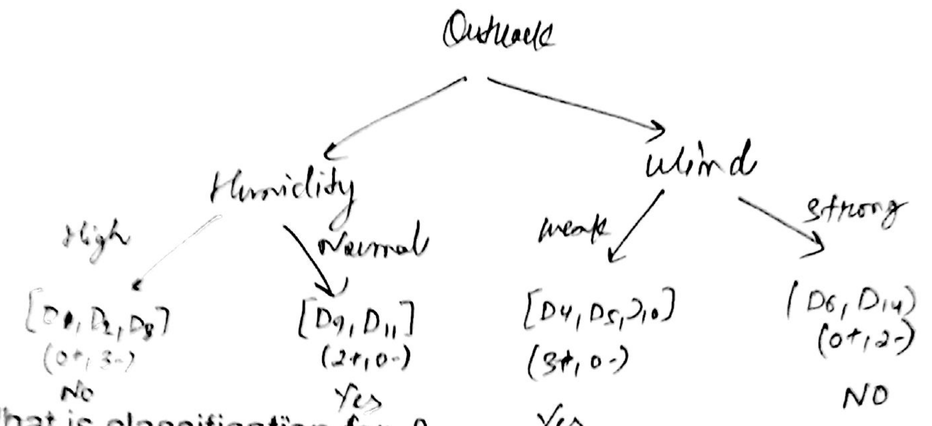
which attribute to test here?

S_{sunny} = [D1, D2, D8, D9, D11]

Gain(Sunny, Humidity) $= 0.970 - \frac{3}{5}(0.0) - \frac{2}{5}(0.0) = 0.970$

Gain(Sunny, Temperature) $= 0.970 - \frac{2}{5}(0.0) - \frac{2}{5}(1.0) = 0.570$

Gain(Sunny, Wind) $= 0.970 - \frac{2}{5}(1.0) - \frac{3}{5}(1.0) = 0.019$



Classification Stops satisfied as all classes segregated

What is classification for ?

(Overlook, Humidity, Wind) = (Rain, High, Weak)

ALERT: (missing value) tell me about Temperature? → Yes

DBSCAN ALGO →

- Partitioning methods and hierarchical clustering are suitable for finding spherical - shaped clusters.
- They work well only for compact and well separated clusters.
- Severely affected by presence of Noise and Outliers.
- No. of clusters need to be Known.

Real-life data -

- Clusters of arbitrary shape (oval, linear, S)
- Many outliers and noise.

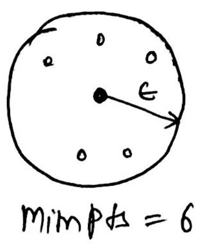
Solution → Density -based.

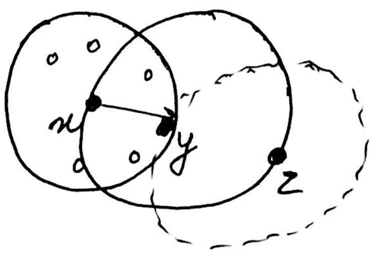
Two Types →

- Connectivity -based : DBSCAN, HDBSCAN, OPTICS
- Density -function based : DENCLUE

DBSCAN

- Goal is to identify dense regions, which can be measured by a no. of objects close to a given point.
- EPSILON "eps" and MIN pts ("minPts")
- "eps" defines the radius of neighbourhood around a point x.
- "minpts" is min. neighbour within "eps" radius.

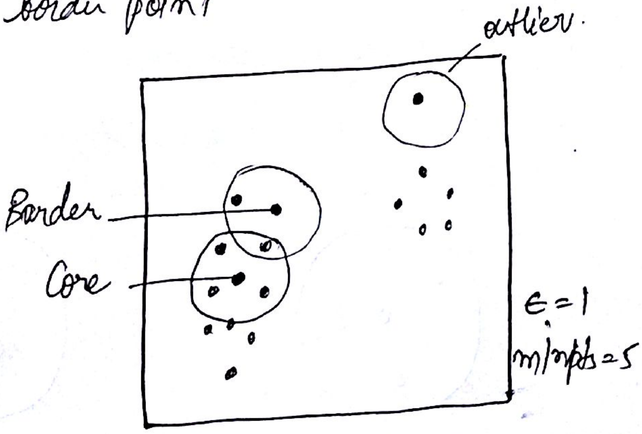




Scanned by CamScanner

DBSCAN or density based Algo

- Density - no. of pts within radius (eps)
- Core point - it has more than a specified no. of points within Eps
  - these are points that are at the interior of a cluster.
- Border point - has fewer points than "Minpts", but neighbour of core pts.
- Noise point - not a core, not a border point



Advantage

- Shape not issue
- No of cluster not required

Marginal Probability → Discrete.

Disease X: 0, 1

Y Has Symptoms: 0, 1

Y (Symptoms) \ X (disease)	0	1
0	0.5	0.1
1	0.1	0.3

$\Sigma = 1$

$\sum_x P(y) = 1$

Joint Probability :

$$P(X=1, Y=1) = 0.3$$

$$\sum_{x,y} P(X=x, Y=y) = 1$$

Marginal Probability :

$$P(y=1) = 0.1 + 0.3 = 0.4$$

$$P(X=0) = 0.5 + 0.1 = 0.6$$

$$\Rightarrow \boxed{P(X=x) = \sum_y P(X=x, Y=y)}$$

Given $P(A=1) = 0.65$, $P(B=1) = 0.77$, find $P(A=1, C=0)$

$$\sum_B P(A=1, B, C=0) \Rightarrow \sum_B P(C=0 / A=1, B) \times P(A=1, B)$$

$$\Rightarrow \sum_B P(C=0/A=1, B) \times P(A=1) \times P(B)$$

$$= P(C=0/A=1, B=0) \times P(A=1) \times P(B=0) + P(C=0/A=1, B=1) \times P(A=1) \times P(B=$$

$$= 0.20 \times 0.65 \times 0.23 + 0.75 \times 0.65 \times 0.77$$

$$= 0.405$$

$P(A=0, C=0)$

$\Rightarrow \sum_B P(A=0, B, C=0)$

$= \sum_B P(C=0 | A=0, B) \times P(A=0) \times P(B)$

$= P(C=0 | A=0, B=0) \times P(A=0) \times P(B=0) + P(C=0 | A=0, B=1) \times P(A=0) \times P(B=1$

$= 0.90 \times 0.35 \times 0.23 + 0.01 \times 0.35 \times 0.77$

$= 0.075$

~~P(A=1 | C=0) = P(~~

$P(A=1, C=0) = P(A=1 | C=0) \times P(C=0)$

$\Rightarrow P(A=1 | C=0) = \dfrac{P(A=1, C=0)}{\sum_A P(C=0, A=A)}$

$= \dfrac{0.405}{0.075 + 0.405} = 0.843$

$P(A=1, B, C=0) = P(C=0 | A=1, B) \times P(A=1) \times P(B)$

$= P(C=0 | A=1, B=0) \times P(A=1) \times P(B=0)$

$P(C=0 | A=1, B=1) \times P(A=1) \times P(B=1)$

$= 0.20 \times 0.65 \times 0.23 + 0.75 \times 0.65 \times 0.77$

$= 405$

Example →

$P(R=1) = 0.2$

$P(S=1) = 0.1$

$P(J=1 \mid R=1) = 1$

$P(J=1 \mid R=0) = 0.2$

$P(T=1 \mid R=1, S=0) = 1$

$P(T=1 \mid R=1, S=1) = 1$

$P(T=1 \mid R=0, S=1) = 0.9$

$P(T=1, R=0, S=0) = 0$

$$P(S=1 \mid T=1) = \frac{P(S=1, T=1)}{P(T=1)}$$

$$= \frac{\sum_{JR} P(S=1, J, R, T=1)}{\sum_{JRS} P(T=1, J, R, S)}$$

$$= \frac{\sum_{JR} P(J \mid R) P(T=1 \mid R, S=1) P(R) P(S=1)}{\sum_{JRS} P(J \mid R) P(T=1 \mid R, S) P(R) P(S)}$$

$$= \frac{\sum_{JR} P(T=1 \mid R, S=1) P(R) P(S=1)}{\sum_{JRS} P(T=1 \mid R, S) P(R) P(S)}$$

$$= \frac{0.9 \times 0.8 \times 0.1 + 1 \times 0.2 \times 0.1}{0.9 \times 0.8 \times 0.1 + 0.2 \times 1 \times 0.1 + 0 + 1 \times 0.2 \times 0.9} = 0.3382$$

Find $P(S=1 \mid T=1) = \frac{P(S=1, T=1)}{P(S=1)}$

We have,

$\cancel{P(T, J, R, S) = P(T \mid J, R, S) \, P(J \mid R) \, P(R) \, P(S)}$

$$P(S=1 \mid T=1) = \frac{P(S=1, T=1)}{P(T=1)} = \frac{0.092}{0.272} = 0.338$$

$$P(S=1, T=1) = P(T, R, S)$$
$$= P(T=1 \mid R=0, S=1) \times P(R=0) \times P(S=1) +$$
$$P(T=1 \mid R=1, S=1) \times P(R=1) \times P(S=1)$$
$$= 0.9 \times 0.8 \times 0.1 + 1 \times 0.2 \times 0.1$$
$$= 0.072 + 0.02 = 0.092$$

$$P(T) = P(T \mid R, S)$$
$$= P(T=1 \mid R=0, S=0) \times P(R=0) \times P(S=0) +$$
$$P(T=1 \mid R=1, S=0) \times P(R=1) \times P(S=0) +$$
$$P(T=1 \mid R=0, S=1) \times P(R=0) \times P(S=1) +$$
$$P(T=1 \mid R=1, S=1) \times P(R=1) \times P(S=1)$$
$$0 \times 0.8 \times 0.9 + 1 \times 0.2 \times 0.9 + 0.9 \times 0.8 \times 0.1 + 1 \times 0.2 \times 0.1$$
$$= 0.18 + 0.072 + 0.02 = 0.272$$