$x \in \mathbb{R}^d$

$\phi \in \mathbb{R}^p \quad p >> d$

feature map

x – attribute

transformed x – features.

Linear Regression.

$$\theta = \theta + \alpha \sum_{i=1}^{n} (y^{(i)} - h_\theta(x^{(i)}))x^{(i)}$$

$$= \theta + \alpha \sum_{i=1}^{n} (y^{(i)} - \theta^T x^{(i)})x^{(i)}$$

ϕ – $\left.\begin{matrix} x \\ \in \\ \mathbb{R}^d \end{matrix}\right\} \to x \in \mathbb{R}^p$

Kernalized linear regression

$$= \theta + \alpha \sum_{i=1}^{n} (y^{(i)} - \theta \phi(x^{(i)})) \cdot \phi(x^{(i)})$$

$$\phi(x) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_1^2 \\ x_1 x_2 \\ x_1 x_3 \\ \vdots \\ x_2 x_1 \\ \vdots \\ x_1^3 \\ x_1 x_2 \\ \vdots \\ \vdots \end{bmatrix}$$

monomial of x with degree 3.

$x \in \mathbb{R}^d$

$\in \mathbb{R}^p$

$d = 1000$

$\phi(x)$ = monomial degree of 3

$p \simeq O(d^3)$

Kernel:-

$$K(x, y) = \langle \phi(x), \phi(y) \rangle$$

Dot product

Kernel func

$x = (x_1, x_2, x_3)$ } Two vectors
$y = (y_1, y_2, y_3)$

$$\phi(x) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ x_1 x_3 \\ x_2 x_1 \\ x_2 x_2 \\ x_2 x_3 \\ x_3 x_1 \\ x_3 x_2 \\ x_3 x_3 \end{bmatrix}$$

$x = [1, 2, 3]$

$y = [4, 5, 6]$

$$\phi(x) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 4 \\ 6 \\ 3 \\ 6 \\ 9 \end{bmatrix} \quad \phi(y) = \begin{bmatrix} 16 \\ 20 \\ 24 \\ 20 \\ 25 \\ 30 \\ 24 \\ 30 \\ 36 \end{bmatrix}$$

$\phi(x) \cdot \phi(y) = 16 + 40 + 72 + 40 + 100 + 144 + 72 + 180 + 324$

$= 1024$

$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0$

12/03/25 L36

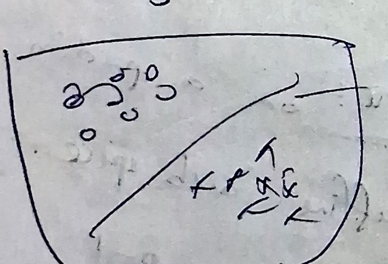
Class $y = \{-1, +1\}$

$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p > 0$, if $y = 1$

$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p < 0$ if $y = -1$

$y_i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_p x_p) > 0$

• Sensitivity



Maximal margin hyperplane/ optimal separating hyperplane

i) d_1, d_2, d_3 d

ii) $\min(d_1, d_2, d_3 \dots d_n)$

max $\begin{cases} M_1 \\ M_2 \\ M_3 \end{cases} \rightarrow$ Best margin

SVM:

1) Maximal Margin Classifier

Maximize M. $\rightarrow$ ①

$\beta_0, \beta_1, \beta_2 \dots \beta_p, M$

Subject to:

$$\sum_{j=1}^{p} \beta_j^2 = 1 \rightarrow ②$$

$$y^{(i)}(\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \dots \beta_p x_p^{(i)}) \geq M \rightarrow ③$$

for all $i = 1, 2, \dots, n$

Constrained optimization problem.

13/03/25 L37

$$d = \frac{|\beta_0 + \beta_1 x_1 + \beta_2 x_2|}{\sqrt{\beta_1^2 + \beta_2^2}}$$

Three lines/hyperplanes.

$2x_1 + 3x_2 - 5 = 0$

$-x_1 + 4x_2 + 7 = 0$

$5x_1 + 12x_2 + 10 = 0$

x_1	x_2	y
3	4	+1
2	3	+1
1	1	-1
-2	1	-1

Line 1 :-

P1 $= \dfrac{|(-5) + 2\cdot 3 + 3\cdot 4|}{\sqrt{2^2 + 3^2}}$

$= \dfrac{|(-5) + 6 + 12|}{\sqrt{4+9}}$

$= \dfrac{13}{\sqrt{13}}$

P2 $\Rightarrow \dfrac{|(-5) + 2\cdot 2 + 3\cdot 3|}{\sqrt{13}}$

$= \dfrac{|-5+4+9|}{\sqrt{13}} = \dfrac{8}{\sqrt{13}}$

P3

$\dfrac{|(-5) + 2\cdot 1 + 3\cdot 1|}{\sqrt{13}} = \dfrac{0}{\sqrt{13}} = 0$

P4 $\dfrac{|(-5) + 2\cdot(-2) + 3(1)|}{\sqrt{13}} = \dfrac{6}{\sqrt{13}}$

Line 2 :-

P1 $\dfrac{|7 + (-1)\cdot 3 + 4\cdot 4|}{\sqrt{(-1)^2 + (4)^2}}$

$= \dfrac{7 - 3 + 16}{\sqrt{17}} = \dfrac{20}{\sqrt{17}}$

P2 $\dfrac{|7 + (-1)\cdot 2 + 4\cdot 3|}{\sqrt{17}}$

$= \dfrac{17}{\sqrt{17}}$

P3 $\dfrac{|7 + (-1)\cdot 1 + 4(1)|}{\sqrt{17}}$

$= \dfrac{10}{\sqrt{17}}$

P4 $\dfrac{|7 + (-1)\cdot(-2) + 4(1)|}{\sqrt{17}}$

$= 13/\sqrt{17}$

Line 3

P1 $\dfrac{|10 + 5(3) + (-12)\cdot(4)|}{\sqrt{5^2 + (-12)^2}}$

$= \dfrac{|10 + 15 + (-48)|}{\sqrt{25 + 144}} = \dfrac{23}{\sqrt{169}}$

P2 $\dfrac{|10 + 5\cdot 2 + (-12)\cdot 3|}{\sqrt{169}} =$

$= 16/\sqrt{169}$

P3 $\dfrac{|10 + 5\cdot 1 + (-12)\cdot 1|}{\sqrt{169}} = \dfrac{3}{\sqrt{169}}$

P4 $\dfrac{|10 + 5(-2) + (-12)\cdot 1|}{\sqrt{169}} = \dfrac{12}{\sqrt{169}}$

Regression Metrics :-

1) $R^2 = 1 - \frac{RSS}{TSS}$

$$= 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}$$

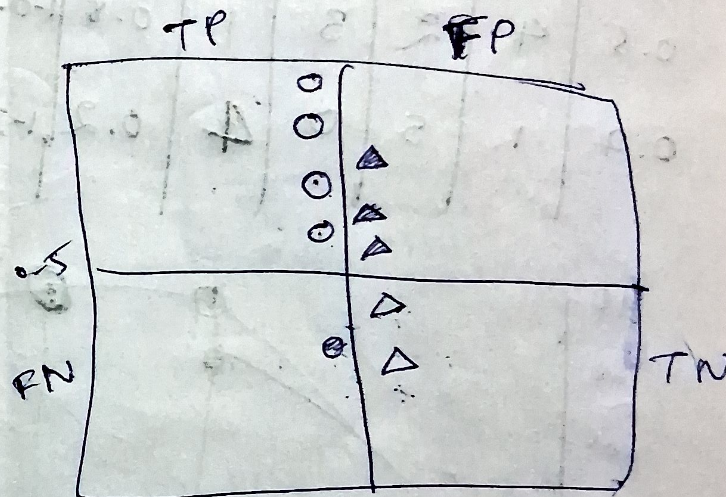
$$R^2 = \frac{Var(mean) - Var(model)}{Var(mean)}$$

* Variance explained by models by data,

21/08/25

L42

1) Accuracy
2) Precision
3) Recall / sensitivity / True positive recall
4) Specificity
5) False positive rate (1 - specificity)
6) F1 - score
7) ROC plot
8) AUC



obser	Actual label	Predicted score	Predicted Label
1	1	0.85	1
2	0	0.60	1 (FP)
3	1	0.70	1
4	1	0.40	0 (FN)
5	0	0.55	1 (FP)
6	1	0.50	1
7	0	0.65	1 (FP)
8	0	0.35	0
9	1	0.60	1
10	0	0.20	0

1) Accuracy $= \frac{4+2}{10} = 0.6$

2) Precision $= \frac{4}{7} = 0.5714$

3) Recall / sensitivity / TPR $= \frac{4}{4+1} = \frac{4}{5} = 0.8$

4) Specificity $= \frac{2}{2+3} = 0.4$

5) False positive rate $= 1 - 0.4$
$= 0.6$

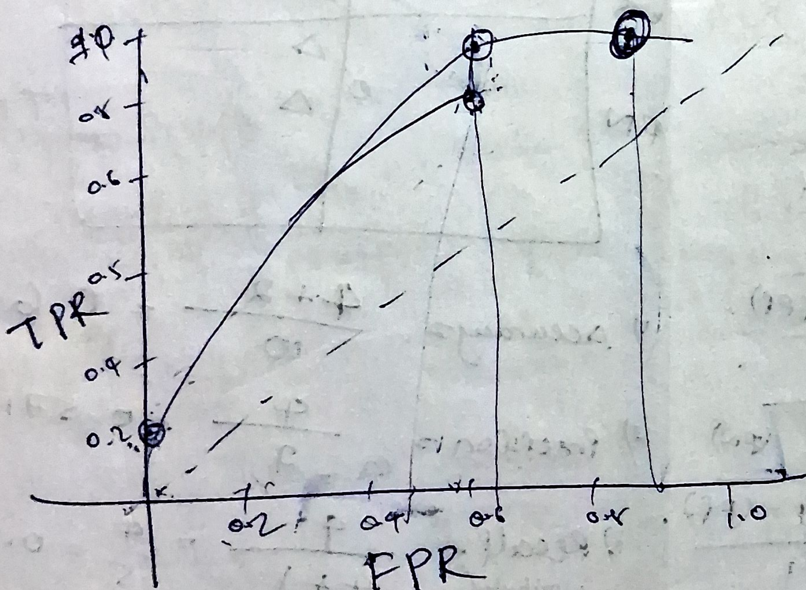
6) F1 score $= 2 \cdot \frac{(0.57) \cdot (0.8)}{(0.57 + 0.8)}$
$= 0.6656934$

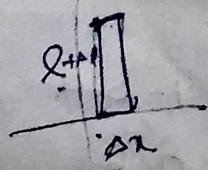
7) ROC :-

Threshold.

0.7
0.5
0.4
0.2

Threshold	TP	TN	FP	FN	TPR	FPR
0.2	5	1	4	0	1	1-0.2 = 0.8
0.4	5	2	3	0	1	1-0.4 = 0.6
0.5	4	2	3	1	0.8	1-0.4 = 0.6
0.7	1	5	0	4	0.2	1-1 = 0





$$AUC = \sum_{i=1}^{4} \frac{(FPR_i - FPR_{i-1}) * (TPR_i + TPR_{i-1})}{2}$$

$$= \sum \frac{(0.6-0) * (0.8+0.2)}{2}$$

$$= 0.3 + \frac{(0.8-0.6) * (1+1)}{2}$$

$$= 0.3 + 0.2$$

$$AUC = \boxed{0.5}$$

25/03/25 L43

Unsupervised Learning:

1) Visualize data, X
2). Generating lower dimensional data of X.
3) Grouping or forming categories on data X.
4). Grouping on data X's features.

Principal Component Analysis (PCA)

Both 1 & 2 challenges

- visualize data
- Primary goal is to reduce dimensionality of data.

$$\{x^{(i)}\}_{i=1}^{n}, \quad x \in \mathbb{R}^d$$

$$x = \underbrace{\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_d \end{bmatrix}}_{d} \Big\} n \quad n \times d$$

22

$$\underbrace{\begin{bmatrix} | & | & | \end{bmatrix}}_{k} \Big\} n \quad n \times k$$

$k < d$

$$= \underset{u}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} \cdot u)^2$$

$$= \underset{u}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \left((x^{(i)T} \cdot u)(x^{(i)T} \cdot u) \right)$$

$$= \underset{u}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \left(u^T x^{(i)} \cdot x^{(i)T} \cdot u \right)$$

$$= \underset{u}{\arg\max} \frac{1}{n} \, u^T \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T} \right)}_{\text{covariance matrix}} u$$

Steps of PCA :

1) Standardize each feature separately
2) Compute $X \cdot X^T$ - covariance matrix
3) Compute the eigen values & eigen vectors of covariance matrix.
4) Decide the % needed of variance explained. and choose 'K' principal components.
5) Transform 'x' into new 'z' space of principal compent.
6) Use the 'z' to train a model.

Example:

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 3 & 2 \\ 1 & 3 \end{bmatrix}$$

1.) Standardize

$$\text{mean } x_1 = \frac{2+0+3+1}{4} = \frac{\cancel{6}^{3}}{\cancel{4}_{2}} = 1.5$$

$$\text{mean } x_2 = \frac{0+1+2+3}{4} = 1.5$$

$$\text{Std} = \sqrt{\frac{(2-1.5)^2+(0-1.5)^2+(3-1.5)^2+(1-1.5)^2}{4-1}}$$

$$= \sqrt{\frac{(0.5)^2+(-1.5)^2+(1.5)^2+(-0.5)^2}{3}}$$

$$= \sqrt{(0.25+2.25+2.25+0.25)/3}$$

$$= \sqrt{5/3} = 1.29 = 1.3$$

$$X = \begin{bmatrix} 0.38 & -1.15 \\ -1.15 & -0.38 \\ 1.15 & 0.38 \\ -0.38 & 1.15 \end{bmatrix}$$

$$\frac{2-1.5}{1.3}$$

Example:

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 3 & 2 \\ 1 & 3 \end{bmatrix}$$

1.) standardize

$$\text{mean } x_1 = \frac{2+0+3+1}{4} = \frac{\cancel{6}^{3}}{\cancel{4}_{2}} = 1.5$$

$$\text{mean } x_2 = \frac{0+1+2+3}{4} = 1.5$$

$$\text{Std} = \sqrt{\frac{(2-1.5)^2+(0-1.5)^2+(3-1.5)^2+(1-1.5)^2}{4-1}}$$

$$= \sqrt{\frac{(0.5)^2+(-1.5)^2+(1.5)^2+(-0.5)^2}{3}}$$

$$= \sqrt{(0.25+2.25+2.25+0.25)/3}$$

$$= \sqrt{5/3} = 1.29 = 1.3$$

$$X = \begin{bmatrix} 0.38 & -1.15 \\ -1.15 & -0.38 \\ 1.15 & 0.38 \\ -0.88 & 1.15 \end{bmatrix}$$

$$\frac{2-1.5}{1.3}$$

2) Covariance matrix

$$= \frac{1}{n-1} X \cdot X^T$$

$$= \frac{1}{n-1} X^T \cdot X$$

$$= \begin{bmatrix} 0.38 & -1.15 & 1.15 & -0.38 \\ -1.15 & -0.38 & 0.38 & 1.15 \end{bmatrix} \cdot \begin{bmatrix} 0.38 & -1.15 \\ -1.15 & -0.38 \\ 1.15 & 0.38 \\ -0.38 & 1.15 \end{bmatrix}$$

$$= \begin{matrix} (0.38)^2 + (-1.15)^2 + (1.15)^2 + (-0.38)^2 & (0.38)(-1.15) + (-1.15) \times (-0.38) + (1.15)(0.38) + (-0.38)(1.15) \\ 0.14 + 1.32 + 1.32 + 0.14 & -0.44 + 0.44 + 0.44 + (-0.44) \\ 2.92 & 0.88 \end{matrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2.92 & 0.88 \\ 0.88 & 2.92 \end{bmatrix}$$

$$= \begin{bmatrix} 0.97 & 0.29 \\ 0.29 & 0.97 \end{bmatrix}$$

3) Compute Eigen value

$$(C - \lambda I) = 0$$

$$\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.97-\lambda & 0.29 \\ 0.29 & 0.97-\lambda \end{bmatrix} = 0$$

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$(0.97-\lambda)^2 - (0.29)^2 = 0$$

$$\lambda^2 + (0.97)^2 - 2(0.97)\lambda - 0.0841 = 0$$

$$\lambda^2 + 0.9409 - 1.94\lambda - 0.0841$$

$$\lambda^2 - 1.94\lambda + 0.857 = 0$$

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4.a.c}}{2a}$$

$$= \frac{+1.94 \pm \sqrt{(-1.94)^2 - 4.1.0.857}}{2}$$

$$= \frac{+1.94 \pm \sqrt{3.76 - 3.43}}{2} = \frac{-1.94 \pm \sqrt{0.33}}{2}$$

$$= \frac{+1.94 \pm 0.57}{2}$$

$\lambda_2 = 0.68$ (35%) $\lambda_1 = 1.25$ (65%)

Total variance = 1.93

Eigen values:

$\lambda_1 = 1.25$ (65% variance explained)

$$\begin{bmatrix} 0.97 - 1.25 & 0.29 \\ 0.29 & 0.97 - 1.25 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -0.28 & 0.29 \\ 0.29 & -0.28 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$-0.28 v_1 + 0.29 v_2 = 0$$

$$0.28 v_1 = 0.29 v_2$$

$$v_2 = \frac{0.28}{0.29} v_1$$

let $v_1 = 1$ $v_2 = 0.96$

$\Rightarrow \begin{bmatrix} 1 \\ 0.96 \end{bmatrix}$

Normalize.

$\sqrt{(1)^2 + (0.96)^2} = \sqrt{1.9216}$

$= 1.39$

$\Rightarrow \begin{bmatrix} 0.72 \\ 0.69 \end{bmatrix}$

5) Transform

$$X = \begin{bmatrix} 0.38 & -1.15 \\ -1.15 & -0.38 \\ 1.15 & 0.38 \\ -0.38 & 1.15 \end{bmatrix} \qquad \text{PCA vector} = \begin{bmatrix} 0.72 \\ 0.69 \end{bmatrix}$$

$$z = \begin{bmatrix} (0.38)(0.72) + (-1.15)(0.69) \\ (-1.15)(0.72) + (-0.38)(0.69) \\ (1.15)(0.72) + (0.38)(0.69) \\ (-0.38)(0.72) + (1.15)(0.69) \end{bmatrix}$$

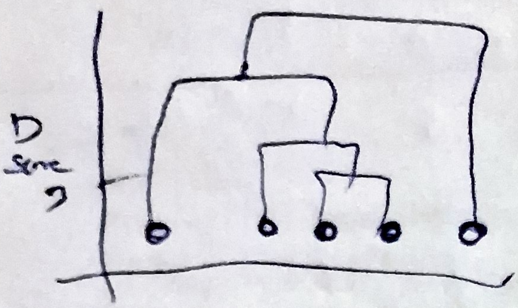
$$z = \begin{bmatrix} 0.27 + (-0.79) \\ -0.83 + (-0.26) \\ 0.83 + 0.26 \\ -0.27 + 0.79 \end{bmatrix}$$

$$Z = \begin{bmatrix} -0.52 \\ -1.09 \\ 1.09 \\ 1.06 \end{bmatrix}$$

02/04/25 L49

Hierarchical Clustering.

- Dendogram/
- Bottom-up tree.



D- Dissimilarity/ Distance

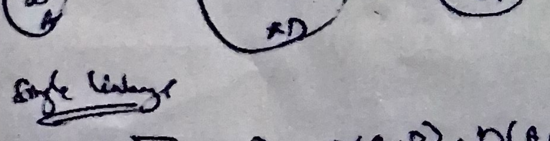
* Clusters — Horizontal cut line.

Linkage :-

similarity b/w cluster & cluster. (or) observation & cluster

- single Linkage
  min(d_1, d_2, d_3)

- Complete Linkage
  max(d_1, d_2, d_3)



Single Linkage

$d_1 = \min[D(A,B),\ D(A,C),\ D(A,D)]$
$= 8.3$

$d_2 = 7.9$

- min - fuse.

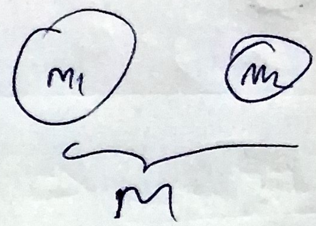
Complete Linkage

min - ✗
max - ✓

$d_1 = \max[D(A,B),\ D(A,C),\ D(A,D)]$

Avg. Linkage :-

- compute the mean of inter cluster distance.



- mean inter cluster distance - Average & mean

Centroid Linkage

similarity / Dissimilarity between
→ Dot product

* Agglomerative / Hierarchical Clustering

	x_1	x_2
A	1	1
B	2	1
C	4	3
D	5	4
E	6	5

- Single Linkage

Euclidean distance

Step 1:- Compute distance matrix

	A	B	C	D	E
A	0	1	3.6	5	6.4
B	1	0	2.8	4.2	5.6
C	3.61	2.8	0	1.41	2.8
D	5	4.2	1.41	0	1.41
E	6.4	5.6	2.8	1.41	0

d_{AB} $\sqrt{0 + (2-1)^2} = 1$

$AC = (3)^2 + (2)^2 = 9 + 4 = \sqrt{13}$

$AD = (4)^2 + (3)^2 = 9 + 16 = \sqrt{25}$

$AE = (5)^2 + (4)^2 = 25 + 16 = \sqrt{41}$

$BD = (3)^2 + (3)^2 = \sqrt{18}$

$BE = (4)^2 + (4)^2 = \sqrt{32}$

$BC = (2)^2 + (2)^2 = \sqrt{8}$

$CD = (1)^2 + (1)^2 = \sqrt{2}$

$CE = (2)^2 + (2)^2 = \sqrt{8}$

$DE = (1)^2 + (1)^2 = \sqrt{2}$

Merges A & B.

New cluster: (AB) C D E

min = 1 distance

	AB	C	D	E
AB	0	2.82	4.24	5.6
C	2.82	0	1.41	2.8
D	4.24	1.41	0	1.41
E	5.6	2.8	1.41	0

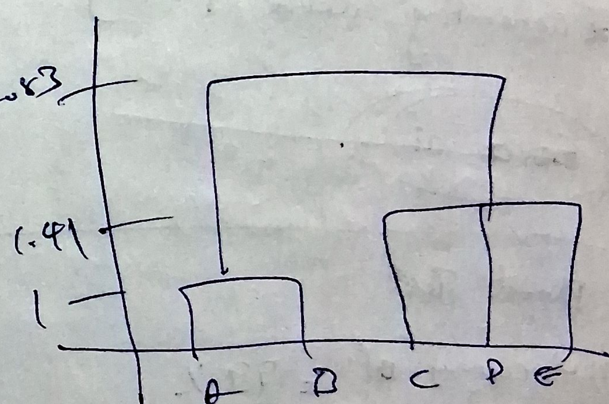
$AC = 3.6$

$BC = 2.8$

$\min(3.6, 2.8)$

$= 2.8$

$\min \begin{cases} AD = 5 \\ BD = 4.2 \end{cases}$



New cluster: (AB) C (DE)

Next cluster: (AB) (CDE)

Final cluster: (ABCDE)

$\underset{y}{\arg\max}\ P(x|y)\,P(y)$

L51

4/04/25

Naive Bayes Classifier

⇒ Spam detection

4% buy this product ⇒ $\begin{bmatrix} \vdots \\ 1.2 \\ 9.8 \\ 0 \\ 7.3 \\ \vdots \end{bmatrix}_{10\times 1}$

Problems/challenges:

⇒ notion of transferring context to a machine

- semantic distance/similarity
- capturing semantic properties

- 4% To buy this product :

$\begin{bmatrix} a \\ aloe \\ \vdots \\ buy \\ \vdots \\ pen \\ product \\ this \\ \vdots \\ whistle \\ \vdots \end{bmatrix}_{50000}$ $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}_{50000\times 1}$

2000

45000

Bernoulli distn / Binomial — only 1 'θ' parameter

Multinomial → vector of 'θ'.

Continuous data → Gaussian distn → 2 parameters

	x_1	x_2	x_3	Class
1	1	0	1	Spam
2	1	1	0	Spam
3	1	1	0	Spam
4	0	0	1	No spam
5	0	0	0	No spam
6	1	1	1	No spam
7	0	1	0	Spam
8	1	0	0	Spam
9	0	1	0	Spam
10	1	0	0	Spam
11	1	1	0	No spam

x_1	x_2	x_3	prob y=1 (spam)	prob y=0 (no spam)
0	0	0	0	1/4
0	0	1	0	1/4
0	1	0	2/7	0
0	1	1	0	0
1	0	0	2/7	0
1	0	1	1/7	0
1	1	0	2/7	1/4
1	1	1	0	1/4

logical AND

$$\theta_{j|y=1} = \frac{\sum_{i=1}^{n} I\{x_j^{(i)}=1 \wedge y^{(i)}=1\}}{\sum_{i=1}^{n} I(y^{(i)}=1)}$$

$$\theta_{j|y=0} = \frac{\sum_{i=1}^{n} I(x_j^{(i)}=1 \wedge y^{(i)}=0)}{\sum_{i=1}^{n} I(y^{(i)}=0)}$$

$$\theta_y = \frac{\sum_{i=1}^{n} I(y^{(i)}=1)}{n}$$

Example

x — y

1) Free win now - spam
2) win a price - spam
3) Hello how are you - not spam
4) Let's win it - Not spam
5) Free lunch ~~to~~ today - Not spam

	word "free"	word "win"	Label
#1	1	1	spam
#2	0	1	spam
#3	0	0	not spam
#4	0	1	not spam
#5	1	0	not spam

1) Calculate prior:

spam $\theta_{y=1} = \frac{2}{5} = 0.4$

no spam $\theta_{y=0} = \frac{3}{5} = 0.6$

$$\theta_{free|y=1} = \frac{1}{2} = 0.5$$

$$\theta_{free|y=1} = 0.5$$

$$\theta_{win|y=1} = \frac{2}{2} = 1$$

$$\theta_{win|y=1} = 1$$

2) Compute Likelihood:

For spam class

$\theta_{free|y=1} = 0.5$

$\theta_{win|y=1} = 1$

For not spam class

$\theta_{free|y=0} = \frac{1}{3}$

$\theta_{win|y=0} = \frac{1}{3}$

Inference Phase

Test message :- Free = yes , win = yes

$P(spam/X) = P(X|spam) P(spam)$

$P(X|spam) = P(X_1, x_2 | spam)$

$= P(x_1|spam) \cdot P(x_2|spam)$

$$P(spam/x) = P(x_1|spam) \cdot P(x_2|spam) \cdot P(spam)$$

$$= P(free = yes|spam) \cdot P(win = yes|spam) \cdot P(spam)$$

$$= \frac{1}{2} \cdot 1 \cdot \frac{2}{5}$$

$$= \frac{1}{5} = \underline{0.2}$$

$$P(Not\ spam/x) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{3}{5}$$

$$= \frac{1}{15} = \underline{0.066}$$

$0.2 > 0.066$

- Implies test message belongs to SPAM class.

09/4/25 (L53)

Bayesian Networks / Probabilistic Graphical Models (PGM) :-

- Directed acyclic graph. (DAG)
- * standard PGM. professor.

$\{r, r, r, b\}$

$p(B_1) =$

r	b
3/4	1/4

B_1

(Marginal)

$P(B_2 | B_1) =$ (Conditional)

B_1 \ B_2	r	b	
r	2/3	1/3	⇒ 1
b	1	0	⇒ 1

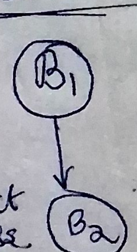
$P(B_1, B_2) = P(B_2, B_1)$

(Joint)

B_1 \ B_2	r	b
r	$\frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$	$\frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$
b	0	0

overall sum upto 1

Simple Graphical Model :-



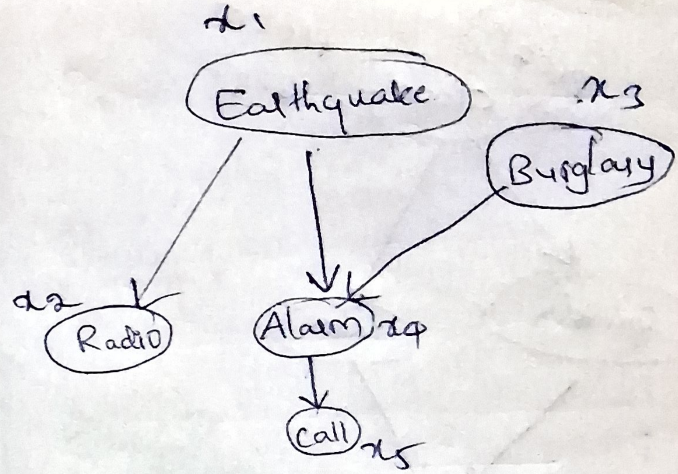
Nodes → Random variables

Edges = direct Influences (causation)

hard to compute

Judea Pearl (Turing award)

$\boxed{B_2 - \text{not caused by } B_1}$



Joint Prob Table — 2^5 rows.

Assumption

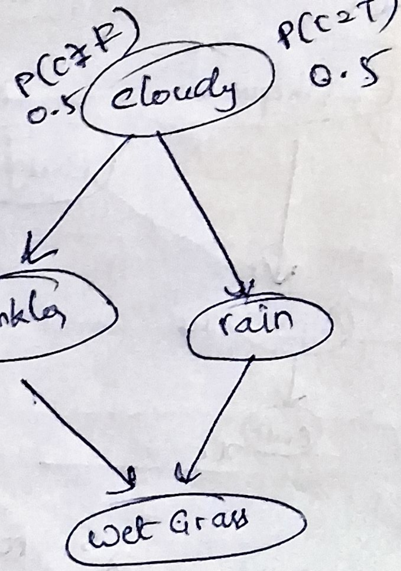
x_i's are independent of x ancestors given x parents.

- Conditional independence.

$P(B, E, A, R, C)$

$= P(B) \cdot P(E|B) \cdot P(A|B, E) \cdot P(R|B, E, A) \cdot P(C|\cancel{B}, \cancel{E}, A, \cancel{R})$

$$\boxed{P(x_1, x_2 \ldots x_d) = \prod_{i=1}^{d} P(x_i | \text{parents}(x_i))}$$



F

P(R|C)

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

P(S|C)

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

P(W|R,S)

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99