$\Theta = \{a, b\}$

- Baye's thm

$$P(X/Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$

Posterior of x — Prior of x

$$P[\theta = Y / d = Y] = \frac{P(d = Y/\theta = Y)\,P(\theta = Y)}{P(d = Y)}$$

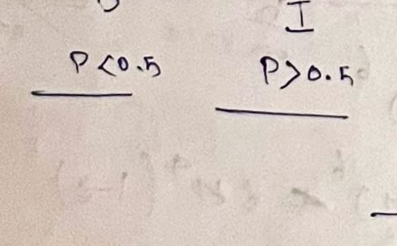
or $$P(\theta/d) = \frac{P(d|\theta)\,P(\theta)}{P(d)}$$

1st Aprill, 24

Logistic Regression

↓

Classifier



$$y = \frac{1}{1 + e^{-(a+bx)}}$$

0 $P \rightarrow$ I

$P < 0.5$ $P > 0.5$

$$P(D) = \frac{1}{1 + e^{-(a+bP)}}$$

Maximum liklihood

Bayesian

4th April. 24

$$P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}$$

eg. $p_H = ?$ $\Rightarrow \theta = \{p_H\}$. $1 \le \theta = p_h \le 1$

$m=10,\ h=4,$
$\mathcal{D}$

$\theta = \theta_1 = 0.2$
$\theta = \theta_2 = 0.8$

$\boxed{\begin{array}{l} \cancel{P(\mathcal{D}|\theta=\theta_1)} \\ \cancel{P(\mathcal{D}|\theta=\theta_2)} \\ \cancel{\text{Max}[P(\mathcal{D}|\theta_j)]} \end{array}}$

Likelihood of θ

$P[\theta|\mathcal{D}] = \frac{P(\mathcal{D}|\theta)\cdot P(\theta)}{P(\mathcal{D})}$

Posterior

Prior

↳ Marginal

So, P(

$$P(\mathcal{D}) = \sum_{i=1}^{m} P\left(\frac{\mathcal{D}}{\theta_i}\right)\cdot P(\theta_i)$$

$$P(\mathcal{D}|\theta) = L(\theta)$$

$\Rightarrow$ Posterior is only dependent on likelihood of θ
to maximize it.

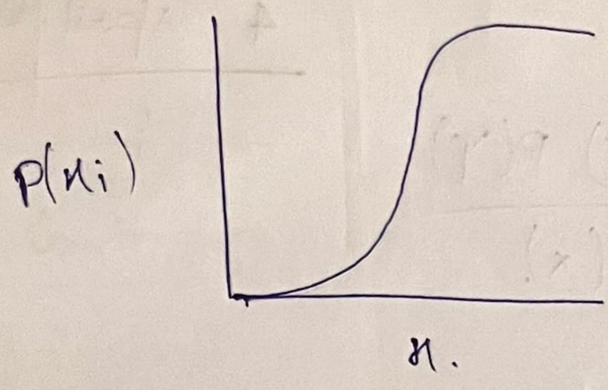
$$\text{Max}[P(\mathcal{D}|\theta)] = {}^{m}C_{h}\, p_h^{\,h} (1-p_h)^{m-h}$$

Given, $\mathcal{D} \Rightarrow m=10,\ h=4$

$$= {}^{10}C_{4}\, p_h^{\,4} (1-p_h)^{6}$$

$\theta = \{p_h\}$

at $p_h = \frac{4}{10}$, it is max.
or $\frac{h}{m}$.



Data

Y	x
0	x_1
1	x_2
y_j	x_i

$$p(x_i) = \frac{1}{1+e^{-(a+bx)}}$$

$y_i = 0$ or 1.

$\theta = \{a, b\}$

Prob. y_i in class 1

$= P(y_i \text{ in class } 1)$

$= P(x_i) = p_i$

$$= \left[\frac{1}{1+e^{-(a+bx_i)}}\right]$$

$P(y_i \text{ in class } 0) = 1 - p_i$

Observⁿ

total Prob.

$\Sigma P(x = x_i) = \prod P(x = x_1)$

$x_1 \to 1$ — $\sigma(x_1)$

$x_2 \to 1$ — $\sigma(x_2)$

$x_3 \to 0$ — $1 - \sigma(x_3)$

total prob

$= \sigma(x_1) \cdot \sigma(x_2) \cdot (1 - \sigma(x_3))$

e.g. BT an example of Bio-pesticides.
(Bacillus thuringiensis)

8th Jan, 24

$$\sigma(x_i) = \frac{1}{1+e^{-(a+b x_i)}}$$

$$P(x = x_i) = \sigma(x_i)^{y_i}\,[1-\sigma(x_i)]^{1-y_i}$$

$$\text{total prob.} = \prod \sigma(x_i)^{y_i}\,[1-\sigma(x_i)]^{1-y_i}$$

↓

$P(D \mid \theta)$

↓

$P[D \mid \theta = \{a, b\}]$

↓

Maximum Likehood

$$\boxed{y_i \in \{0,1\}}$$

Cannot maximize directly because of overflow due to multiplication of fractions.

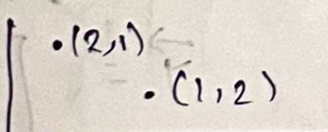
$$L(\theta) = \prod \sigma(x_i)^{y_i}\,(1-\sigma(x_i))^{1-y_i}$$

$\Rightarrow \ln(L(\theta))$ can use to max. it.

Note Gradient Descent or Ascent.

---

ex.



x	y
2	1
1	2
$\bar{x} = \frac{3}{2}$	$\frac{3}{2}$

x	-x
1	2
2	3
$\frac{3}{2}$	$\frac{5}{2}$

$$\frac{\frac{-1}{2}}{(\frac{1}{4}\,\frac{1}{2}\,\frac{1}{4})\cdot\frac{1}{2}}$$

$\text{cov} = (x - x_1)(y - y_1)$

$$\frac{1}{2}\cdot\left(\frac{-1}{2} + \frac{-1}{2}\,\frac{1}{2}\right)$$

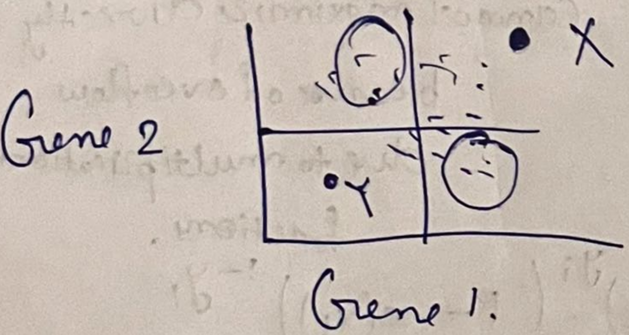
12th April, 24

tillnow

Classification
(supervised learning).
- Random forest.
- Ada boost.
- Support vector machine

Today

Clustering [unsupervised learning]



(features)
Space.

Similarity
↓
distance
$\vec{X}$ $\vec{Y}$
↓
$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

if we have m-dimensional feature space

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

euclidian distance

$$d_{xy} = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{1/2}$$

- Manhattan distance

$$d_{xy} = \sum_{i=1}^{m} |x_i - y_i|$$

- Minkowski distance

$$d_{xy} = \left|(x_i - y_1)^p\right|^{1/p}, \; p \geqslant 1$$

$p = 1$, manhattan

$p = 2$, euclidian.

$$e^{-\frac{(x-\mu)^T \cdot S^{-1}(x-\mu)}{2}}$$

$$d^2_{xy} = \sum_{i=1}^{m} (x_1 - y_i)^2$$

$$= (\vec{x} - \vec{y})^T (\vec{x} - \vec{y})$$

$$= (\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})$$

$$S = \begin{bmatrix} v_1 & 0 \\ 0 & v_2 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} \frac{1}{v_1} & 0 \\ 0 & \frac{1}{v_2} \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$d^2_{xy} = \begin{bmatrix} x_1 - y_1 & x_2 - y_2 \end{bmatrix} \begin{bmatrix} \frac{1}{v} & 0 \\ 0 & \frac{1}{v} \end{bmatrix} \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \end{bmatrix}$$

$$= \frac{1}{v_1}(x_1 - y_1)^2 + \frac{1}{v_2}(x_2 - y_2)^2$$

weightage ↓ $\Rightarrow$ weightage ↑

Let $S = \begin{vmatrix} v_1 & c \\ c & v_2 \end{vmatrix} \Rightarrow S^{-1} = \frac{1}{v_1 v_2 - c^2} \begin{vmatrix} v_2 & -c \\ -c & v_1 \end{vmatrix}$

Note: Mahalonbis distance.

$$d_{x,y}^2 = \left[(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y}) \right]$$

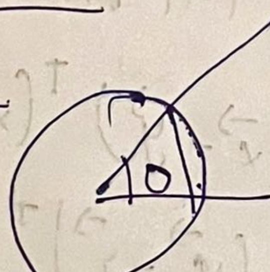
; if $S = I$,

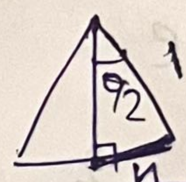
d_{xy} = ~~Euclidean~~

Chord distance

exp2.

exp1





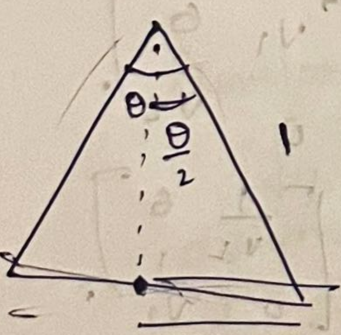
$2 \cos \frac{\theta}{2}$

$\cos\theta = 2\cos^2\frac{\theta}{2} - 1$

$\sqrt{\frac{1+\cos\theta}{2}}$

$d_{xy} = \sqrt{2(1+\cos\theta)}$

$2(1+\theta)$



$C = 2\cos\frac{\theta}{2} = 2 \cdot \sqrt{\frac{1+\cos\theta}{2}} = \sqrt{2(1+\cos}$

$\cos\theta = 2\cos^2\frac{\theta}{2} - 1$

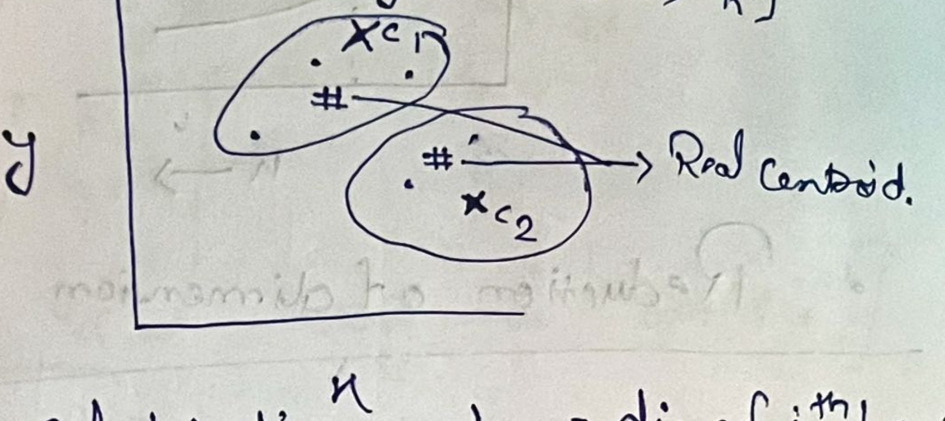
$\cos\frac{\theta}{2} = \sqrt{\frac{1+\cos\theta}{2}}$

19th April, 24

- K-means [1. algorithm first define k] ↳ no. of cluster or subset.

X	Y

2. Seed k centroid randomly. [$C_1, C_2, \ldots, C_K$]



3. Calculate distance. d_{ij} = dis of i^{th} point from j^{th} centroid. $\forall\ i, j$.
4. get $\min(d_{ij})$ for i.
5. based on min, assign i^{th} data to a cluster.
6. Calculate the real centroids of cluster
7. go to 3
8.

- Within cluster Sum of Square:

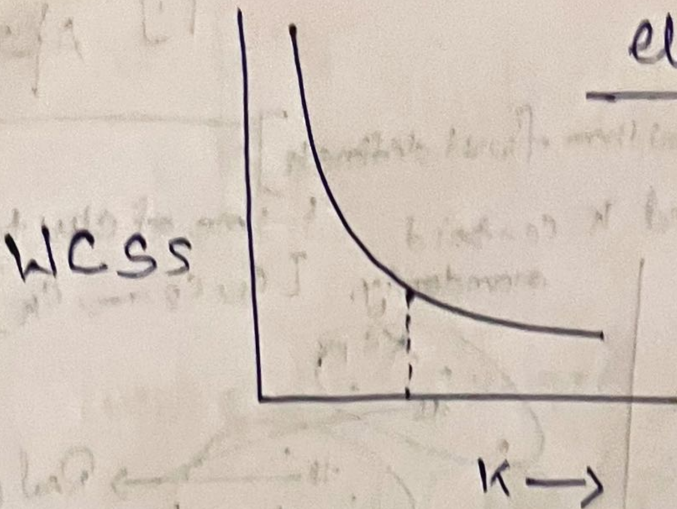
WCSS

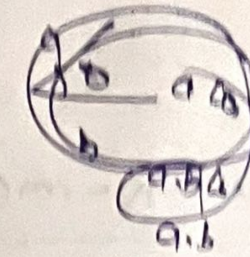
$$WCSS = \sum_{q} \sum_{i,j} \left(d_{ij}^{q} - \bar{d}_{avg}^{q}\right)^2$$

BCSS between

$$BCSS = \sum_{n,m} \left(d_{nm} - \bar{d}_c\right)^2$$

↳ clusters, $m \neq n$

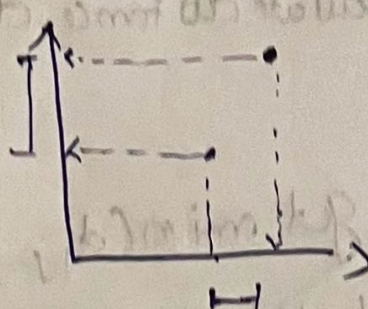




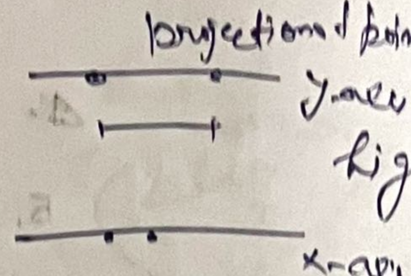
elbowplot

• Reduction of dimension

eg.



2D → 1D



find a vector on which variance of the projection of original data is max.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{d1} \\ x_{12} & & \vdots \\ \vdots & & \vdots \\ x_{1m} & \cdots & x_{dm} \end{bmatrix}_{m \times d}$$

$\vec{u} = d \times 1$.

$$p_1 = \frac{\begin{bmatrix} x_{11} \\ \vdots \\ x_{dm} \end{bmatrix} \cdot \vec{u}}{|\vec{u}|}$$ projection of 1^{st} datapoint

$\therefore$ $p_1 = \frac{1}{|\vec{u}|} [x_{11} \ldots x_{dm}] \cdot \vec{u}$

all proj. $\underset{m \times 1}{P} = \frac{X\vec{u}}{|u|}$; if $|u| = 1$

$$P = X\vec{u} = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}_{m\times 1}.$$

Choose $\vec{u}$; $var(P)$ is max.

$$\vec{u}_{d\times 1} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{bmatrix} \qquad \arg\max\left[var(X\vec{u})\right]$$

$$\sqrt{u_1^2 + u_2^2 + \ldots u_d^2} = 1 \qquad \text{with constraint } |\vec{u}|^2 = 1.$$

Principle Component Analysis

$$Var(\vec{v}) = \frac{1}{m}\vec{v}_c^{\,T} v_c \,, \quad \vec{v} = \begin{vmatrix} v_1 \\ v_2 \end{vmatrix}, \vec{v}_c = \begin{vmatrix} v_1 - \bar{v} \\ v_2 - \bar{v} \end{vmatrix}$$

Centered data.

Illy. for $\vec{u}$.

$$Var(X\vec{u}) = \frac{1}{m}(X\vec{u})_c^T (X\vec{u})_c$$

$$= \frac{1}{m}(X_c\vec{u})^T (X_c\vec{u}) \left[\text{algebrically Same}\right]$$

$X_c \equiv X$ [Already given centred data].

$$= \frac{1}{m}(X\vec{u})^T (X\vec{u}).$$

$$= \frac{1}{m}\vec{u}^T X^T X \vec{u} = u^T S \vec{u}$$

e.g. $\begin{bmatrix} x & y \end{bmatrix}^T \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

$h = ax^2 + 2bxy + cy^2$

So, $\vec{u}^T S \vec{u}$, quadratic form.

Scalar

$\frac{dh}{d\vec{x}} = \nabla h = \begin{bmatrix} \frac{\partial}{\partial x} h \\ \frac{\partial}{\partial y} h \end{bmatrix}$.

vector

$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

$\frac{dh}{dx} \begin{bmatrix} 2ax + 2by \\ 2bx + 2cy \end{bmatrix} = 2 \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

$S\vec{u} = \lambda \vec{u}$

$\vec{u}$ is an eigenvector of S, λ is the eigen vector.

if $\vec{X}_{m \times d}$; $\vec{S}_{d \times d}$

↓

if variable of x are independent . $m > d$.

are 1

then, no. of eigenvector is $\vec{u}_1, \ldots \vec{u}_d$.

Corresponding eign values $\lambda_1, \lambda_2, \ldots, \lambda_d$

$\forall i, \ \lambda_i > 0.$

When, $m < d$.

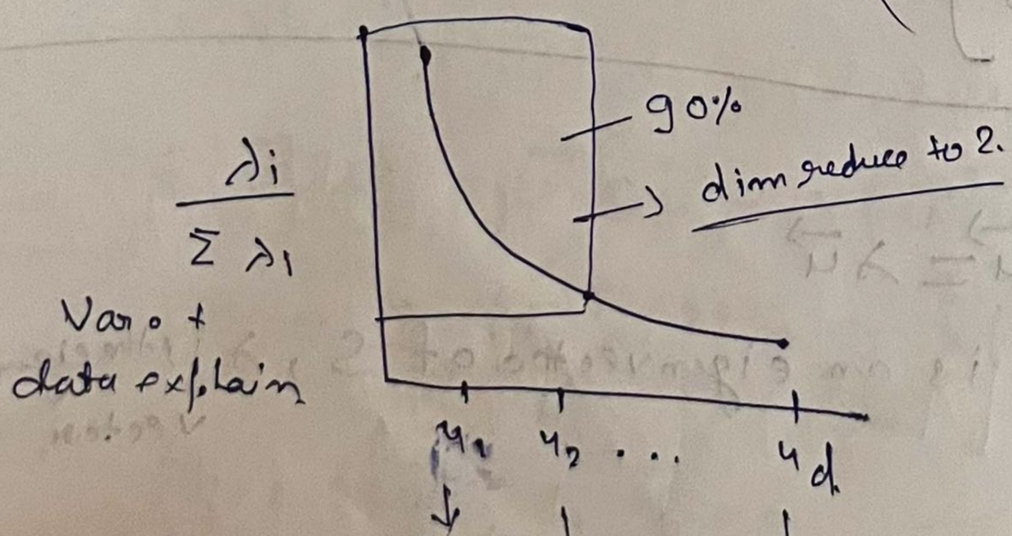
at least one $\lambda = 0$.

Q. $S\vec{u}_i = \lambda_i \vec{u}_i$

Prove it

$Var(X\vec{u}_i) = \lambda_i$

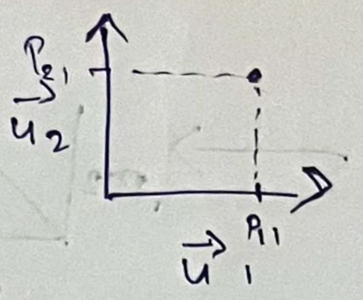
$\left(\Sigma \lambda_j = Var(x) \right)$



$\lambda_1 > \lambda_2 > \ldots > \lambda_d$

$X\vec{u}_1 = P_{c1}$

$X\vec{u}_2 = P_2.$



$$X\begin{bmatrix} | & | \\ \vec{u}_1 & \vec{u}_2 \\ | & | \end{bmatrix} = \tilde{P}$$

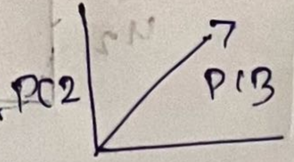
$$\begin{bmatrix} | & | \\ \vec{p}_1 & \vec{p}_2 \\ | & | \end{bmatrix}$$

$$P_1 = \begin{bmatrix} p_{11} \\ \vdots \\ p_{1m} \end{bmatrix} \qquad P_2 = \begin{bmatrix} p_{21} \\ \vdots \\ p_{2m} \end{bmatrix}$$

$$X_{m\times d}\underbrace{\begin{bmatrix} | & | \\ \vec{u}_1 & \vec{u}_2 \\ | & | \end{bmatrix}}_{d\times 2} = \begin{bmatrix} | & | \\ \vec{p}_1 & \vec{p}_2 \\ | & | \end{bmatrix}_{m\times 2.}$$

26th April, 24

$X_{m \times d}$ $\longrightarrow$



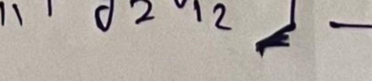
d' dimension $<< d$.

$$\vec{u} = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_d \\ | & | & & | \end{bmatrix}_{d \times d} \rightarrow \text{loading matrix.}$$

$$\vec{X}\vec{U} = P_{m \times d} = \begin{bmatrix} | & | & & | \\ \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_d \\ | & | & & | \end{bmatrix}$$

Eg. $X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$ $\vec{u}_1 = \begin{bmatrix} \vec{v}_{11} \\ \vec{v}_{12} \end{bmatrix}$

$$X u_1 = \begin{bmatrix} x_1 v_{11} + y_1 v_{12} \\ x_2 v_{11} + y_2 v_{12} \end{bmatrix} \begin{matrix} — \text{o} \\ — \times \end{matrix}$$



PCA:- Principal Component Analysis
linear method

Alternative :- tSNE.
uMAP