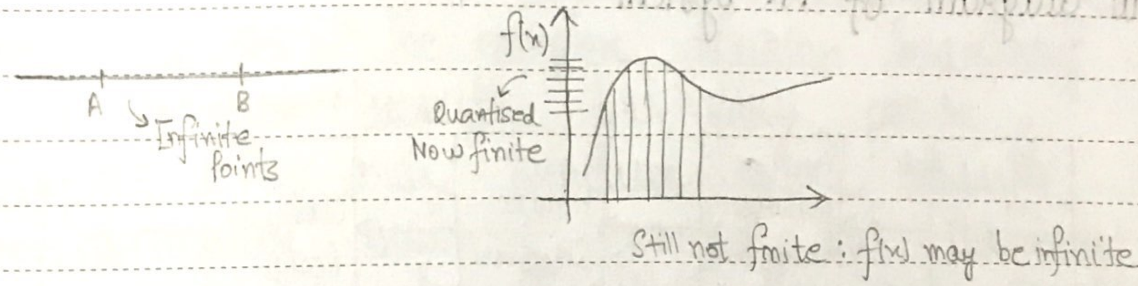
Anant Ashutosh Sharma

2K15/CO/026

Mid: 20
Class: 15
Experiment: 25
End: 40

Pattern Recognition



Capture → Preprocess → Selection & Detection

Feature Vector:

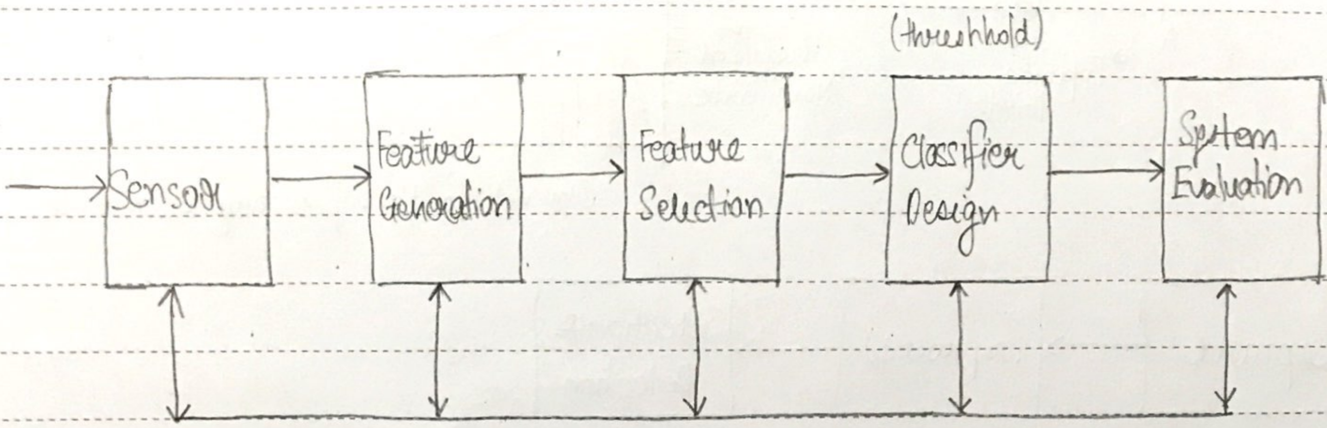
$$V = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \qquad V = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

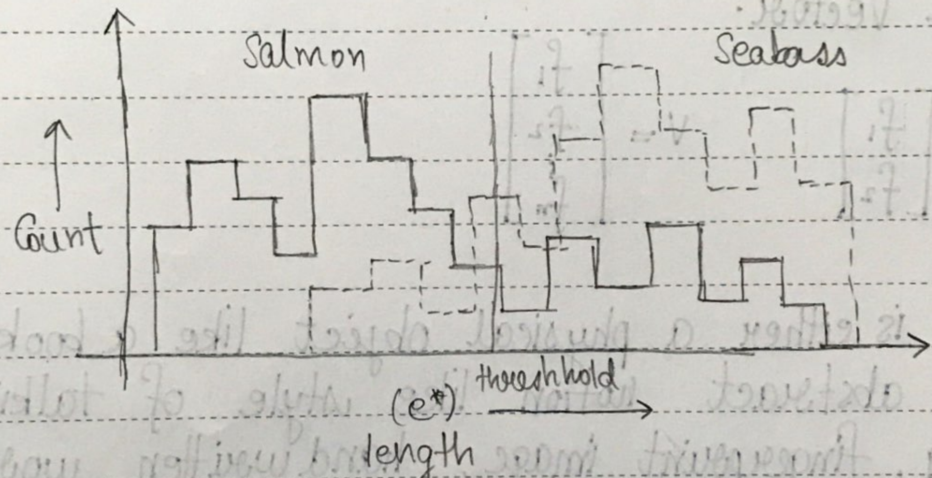
Pattern is either a physical object like a book or chair or an abstract notion like style of talking or writing, fingerprint image, handwritten word etc.

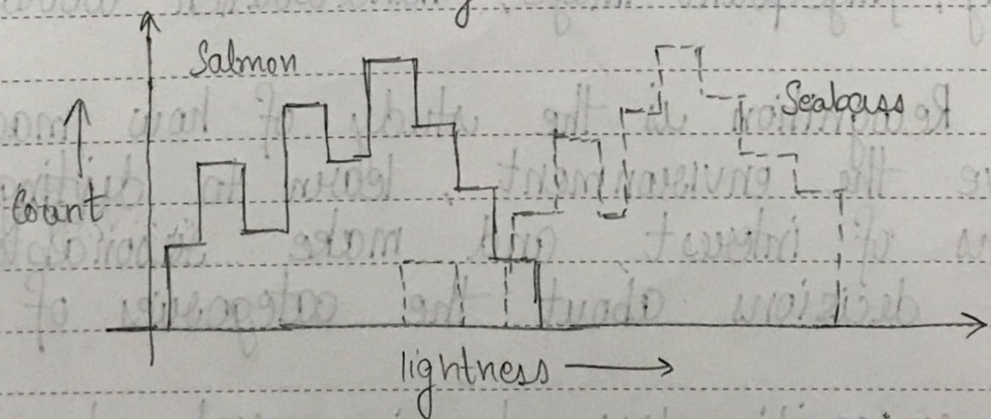
Pattern Recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest, and make reasonable and sound decisions about the categories of patterns.

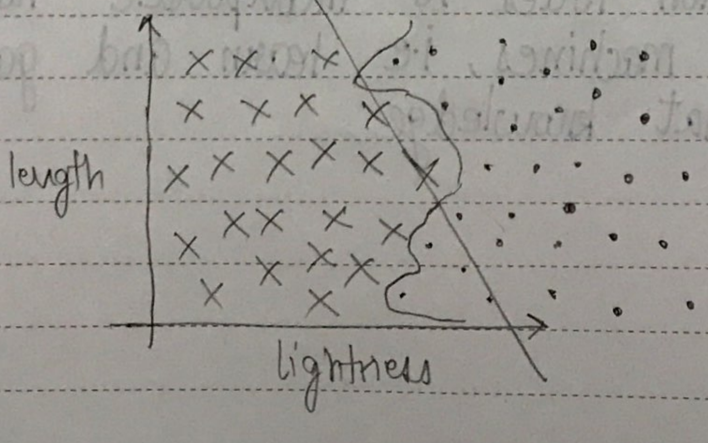
Pattern recognition tries to incorporate human intelligence in machines, i.e. learn and gain knowledge and apply that knowledge

Flow diagram of PR System









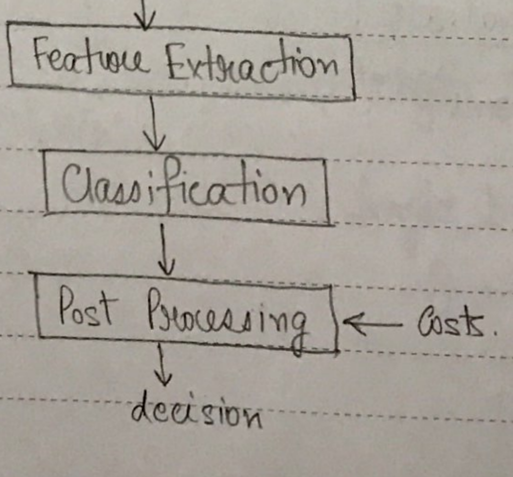
We can consider different decision boundaries to get a proper classifier. The complex decision boundary gives you optimum results but very poor generalization. They may produce good results for a particular training example but unlikely to produce proper classification for new sample. Simple curves give better generalization.

Dimensionality

More features i.e. feature vector with higher dimension may be used to achieve better classification. Although more features may not always produce better results. Sometimes, it even degrades the performance of the classifier if features are not selected properly.

Redundancy

Some features may provide related information, i.e. by increasing the number of features, little addition of information takes place. This results in redundant features.

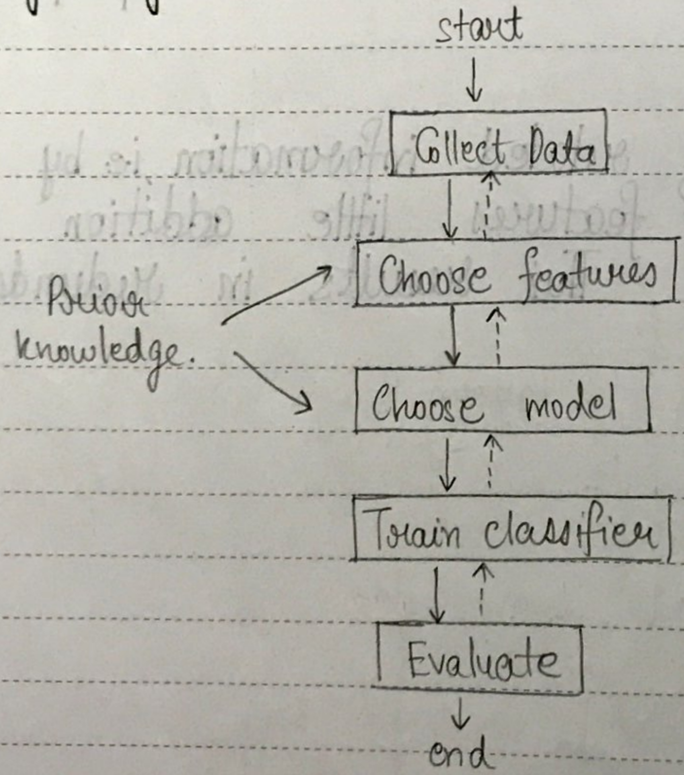


Pattern Recognition System

Feature Extraction:
An ideal feature extractor would yeild a representation that makes the job of a classifier trivial. Conversely, an omnipotent classifier could not need the help of a sophisticated feature extractor.
* It is desireable to have distinguishing features that are invariant to irrelevant transformations of the input

Noise
Any property of the sensed pattern, which is not due to the two underline model but instead because of the randomness in the world.

Design Cycle



Classification

The task of a classifier component is to use the feature vector provided by the feature extractor, to assign the object to a category.

Coloumn Vector $\vec{V} = \begin{bmatrix} \vec{V_1} \\ \vec{V_2} \end{bmatrix}$ $\begin{array}{l} \rightarrow \text{I feature} \\ \rightarrow \text{II feature} \end{array}$

Vector Addition

$$\vec{V} = \begin{bmatrix} \vec{V_1} \\ \vec{V_2} \end{bmatrix} \qquad \vec{W} = \begin{bmatrix} \vec{W_1} \\ \vec{W_2} \end{bmatrix}$$

$$\vec{V} + \vec{W} = \begin{bmatrix} \vec{V_1} + \vec{W_1} \\ \vec{V_2} + \vec{W_2} \end{bmatrix}$$

Scalar Multiplication

$$\vec{V} = \begin{bmatrix} \vec{V_1} \\ \vec{V_2} \end{bmatrix} \qquad 2\vec{V} = \begin{bmatrix} 2V_1 \\ 2V_2 \end{bmatrix}$$

Linear Combination

$c\vec{V} + d\vec{W} = R$ c & d are constants

eg: $Ax = b$

$$\begin{array}{l} 2x - y = 0 \\ -x + 2y = 3 \end{array} \Rightarrow \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$ Coloumn Picture

Inner Product

$\overline{V} \cdot \overline{W} = V_1 W_1 + V_2 W_2$

if $\overline{V} \cdot \overline{W} = 0$

the $\vec{V} \perp \vec{W}$

Length $= \sqrt{\vec{V} \cdot \vec{W}}$

Angle between two

$$\theta = \cos^{-1}\left(\frac{\overline{V} \cdot \overline{W}}{|\overline{V}||\overline{W}|} \right)$$

Matrix Multiplication

$$\begin{bmatrix}2&5\\1&3\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$$

$$= 1\begin{bmatrix}2\\1\end{bmatrix} + 2\begin{bmatrix}5\\3\end{bmatrix}$$

$$= \begin{bmatrix}12\\7\end{bmatrix}$$

$$\begin{bmatrix}a & b & c\end{bmatrix}\begin{bmatrix}c_1 & c_2 & c_3\\ | & | & |\end{bmatrix}$$

→ a×column 1 + b×colomn 2 + c×column 3

Exchange Row

$$\begin{bmatrix}a&b\\c&d\end{bmatrix} \Rightarrow \begin{bmatrix}c&d\\a&b\end{bmatrix}$$

Row1 = 0×Row1 + 1×Row2

Row 2 →

Column Exchange

$$\begin{bmatrix}a&b\\c&d\end{bmatrix} \Rightarrow \begin{bmatrix}b&a\\d&c\end{bmatrix}$$

C1 = 0×C1 + 1×C2

C2 = 0×C2 + 1×C1

Method ①

$$[A]_{m\times n}\,[B]_{n\times p} = [C]_{m\times p}$$

c_{ij} = (i^{th} row of A)(j^{th} column of B)

$$= a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \dots$$

$$= \sum_{k=1}^{n} a_{ik}b_{kj}$$

Method ②

$$[A]\begin{bmatrix}c_1 & c_2 & c_3\\ | & |\ B & |\end{bmatrix} = [C]$$

Matrix A × 1st Coloumn of B = 1st Coloumn of C

2nd

3rd

Method ③

Using Row Vector

$$\begin{bmatrix}\frac{R_1}{R_2}\\ R_3 \ \ A\end{bmatrix}[B] = [C]$$

R_1 × Matrix B = r_1 of C

R_2 ..

R_3 ...

Method ④

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$$

$C_1 = A_1B_1 + A_2B_3$

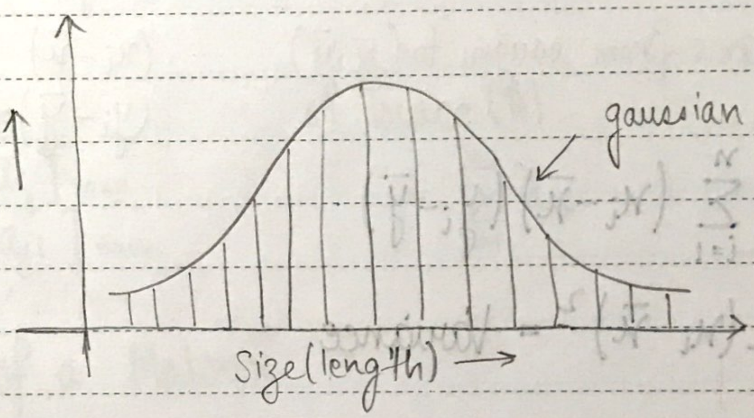
$C_2 = A_1B_2 + A_2B_2$

$C_3 = A_3B_1 + A_1B_3$

$C_4 = A_3B_2 + A_4B_4$

Gaussian PDF is the most widely used distribution.

ALL METHODS of INVERSION.



Single Variable:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$\mu \rightarrow$ Mean

$\sigma^2 \rightarrow$ Variance

Vector

$|X|$

$$p(\underline{x}) = \frac{1}{(\sqrt{2\pi})^n \underbrace{|\Sigma|}_{\text{Mod}}^{1/2}} \exp\left(-\frac{1}{2} \underbrace{\underbrace{(\underline{x}-\underline{\mu})^T}_{1\times n} \underbrace{\Sigma^{-1}}_{n\times n} \underbrace{(x-\mu)}_{n\times 1}}_{\text{Scalar}}\right)$$

$\underline{x}$ _Vector

$\downarrow$

vector $n \times 1$

Σ - Variance Covariance Matrix / Dispersion Matrix

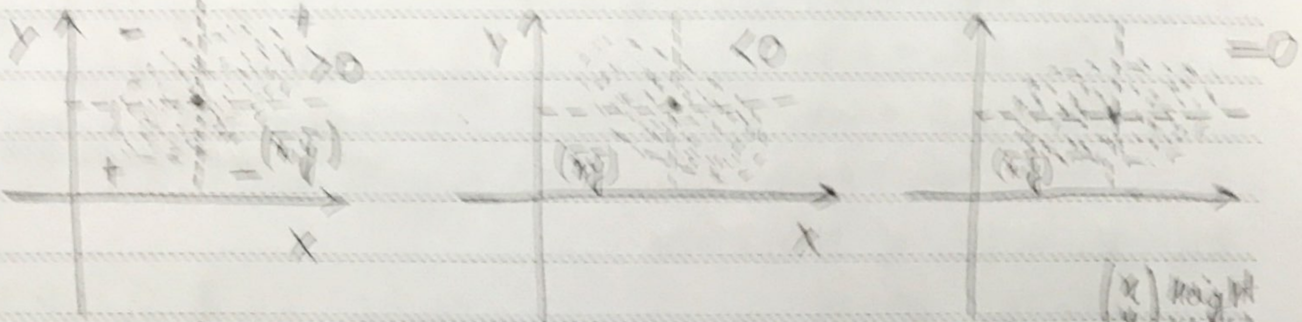
μ - mean vector

$\underline{x} \in \mathbb{R}^n$

$$\int_{\mathbb{R}^n} p(\underline{x})\,dx = 1$$

$p(\underline{x}) \geq 0$

$|\Sigma|$ determinant of variance covariance matrix cannot be 0 or negative, we need to ensure that determinant of this matrix should be strictly positive.



Shifting the axis from origin to $(\bar{x}, \bar{y})$

$(x_i - \bar{x})$
$(y_i - \bar{y})$

$\begin{pmatrix} x \\ y \end{pmatrix}$ height, weight

$\bar{x} \rightarrow$ mean of x
$\bar{y} \rightarrow$ mean of y

$$\underset{Cov(X,Y)}{\text{Covariance}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$Cov(X,X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \text{Variance}$$

for n number of features

$$\text{Variance Covariance Matrix} = \begin{bmatrix} Cov(X_1 X_1) & Cov(X_1 X_2) & \cdots & Cov(X_1 X_n) \\ Cov(X_2 X_1) & Cov(X_2 X_2) & \cdots & Cov(X_2 X_n) \\ \vdots & \vdots & & \vdots \\ Cov(X_n X_1) & Cov(X_n X_2) & \cdots & Cov(X_n X_n) \end{bmatrix}$$

$$\text{Euclidean Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = \sqrt{\sum (x_i - y_i)^2}$$

$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$

$\begin{pmatrix} 160\text{ cm} \\ 70\text{ kg} \end{pmatrix} \begin{pmatrix} 168\text{ cm} \\ 74\text{ kg} \end{pmatrix}$ $\quad$ $\begin{pmatrix} 1600\text{ mm} \\ 70\text{ kg} \end{pmatrix} \begin{pmatrix} 1680\text{ mm} \\ 74\text{ kg} \end{pmatrix}$

$\sqrt{8^2 + 4^2} = \sqrt{80}$ (when in cm) $\quad$ $\sqrt{(80)^2 + 4^2} = \sqrt{6416}$

Matrix Operations

1) Equality

$A = B$

$a_{ij} = b_{ij}$

2) Addition and Subtraction

$A + B = C \qquad A - B = C$

$a_{ij} + b_{ij} = c_{ij} \qquad a_{ij} - b_{ij} = c_{ij}$

3) Scalar Multiple

$K \cdot A = K \begin{bmatrix} & \\ & \end{bmatrix}$

4) Matrix Multiplication

$C_{m \times p} = A_{m \times n} * B_{n \times p}$

5) Transpose

$A = [a_{ij}]_{m \times n}$

$A^T = [a_{ji}]_{n \times m}$

6) Trace (A)

$= \sum_{i=1}^{n} a_{ii}$

Inverse of a Matrix

$A \rightarrow A^{-1}$

Let $A^{-1} = B$

$AB = BA = I$

$AA^{-1} = A^{-1}A = I$

$|A| \neq 0$

A is non singular

and A should be square matrix.

Matrix Inverse Properties

a) If (AB) is invertible

$(A \cdot B)^{-1} = B^{-1}A^{-1}$

b) $(A^{-1})^{-1} = A$

c) If A^n is invertible

$(A^n)^{-1} = (A^{-1})^n = A^{-n}$

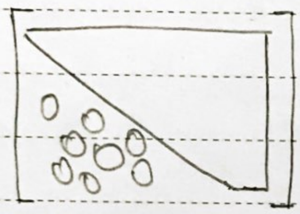
d) $(A^T)^{-1} = (A^{-1})^T$

Find Inverse

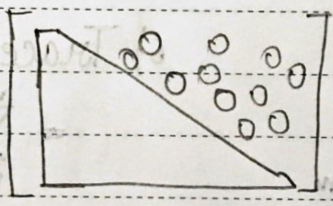
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$A^{-1} = \frac{1}{(ad-bc)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Triangular Matrix



upper Right



lower left

Symmetric Matrix

$A \Rightarrow$ if $a_{ij} = a_{ji}$ $\quad a_{12} = a_{21}$

* (AA^T) & (A^TA) are symmetric

$\Rightarrow$ If A is invertible and symmetric
then (A^{-1}) is also symmetric

$\Rightarrow$ If A is invertible
(AA^T) & (A^TA) are both invertible

Eigen Value & Vector

Let $A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Eigen Vector, associated with eigen value (λ)

$$A \cdot x = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda x$$

$(A - \lambda I)x = 0 \rightarrow$ Characteristic Equation / Polynomial

$|A - \lambda I| = 0$

is singular

$$\begin{vmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{vmatrix} = 0$$

$(\lambda - 3)(\lambda - 2) = 0 \Rightarrow \lambda = 3, 2$

$\lambda = 2$ $\begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$ $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} k$ any scalar value

$\lambda = 3$ $\begin{bmatrix} -2 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$

$-2x_1 + x_2 = 0$ $x = k\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ or $k\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$

$x_1 = \frac{x_2}{2}$

Example:

$$A = \begin{bmatrix} -21 & -9 & 12 \\ 0 & 6 & 0 \\ -24 & -8 & 15 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} -21-\lambda & -9 & 12 \\ 0 & 6-\lambda & 0 \\ -24 & -8 & 15-\lambda \end{bmatrix}$$

$|A - \lambda I| = 0 \Rightarrow -(\lambda - 6)(\lambda - 3)(\lambda + 9) = 0$

$\Rightarrow \lambda = 6, 3, -9$

$\lambda_3 = -9$

$$A + 9I = \begin{bmatrix} -12 & -9 & 12 \\ 0 & 15 & 0 \\ -24 & -8 & 24 \end{bmatrix}$$

Reduce row to get row echelon form

$R_3 = R_3 - 2R_1$
$$\begin{bmatrix} -12 & -9 & 12 \\ 0 & 15 & 0 \\ 0 & 10 & 0 \end{bmatrix}$$

$R_3 = R_3 - \frac{2}{3} R_2$
$$\begin{bmatrix} -12 & -9 & 12 \\ 0 & 15 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_1 = \frac{5}{3} R_1 + R_2$
$$\begin{bmatrix} -20 & 0 & 20 \\ 0 & 15 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_1 = \frac{-1}{20} R_1$ $\quad$ $R_3 = \frac{1}{15} R_3$
$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = 0$$

$V_1 - V_3 = 0$

$V_2 = 0$ $\quad$ Let $V_3 = e_1$

$V_1 = e_1$

$\lambda_3 = -9$

$$V^{(3)} = e_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$\lambda = 6$ $\quad$ $\lambda = 3$

$$V = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \quad V = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}$$

$$A=\begin{bmatrix}5&0&-1\\0&8&0\\-3&0&7\end{bmatrix}$$

$$\begin{bmatrix}5-\lambda&0&-1\\0&8-\lambda&0\\-3&0&7-\lambda\end{bmatrix}$$

$$(5-\lambda)(8-\lambda)(7-\lambda)-3(8-\lambda)=0$$
$$(8-\lambda)\left[(5-\lambda)(7-\lambda)-3\right]=0$$
$$(8-\lambda)\left[\lambda^2-12\lambda+35-3\right]=0$$
$$(8-\lambda)\left[\lambda^2-12\lambda+32\right]=0$$
$$(8-\lambda)(\lambda-8)(\lambda-4)=0$$
$$\lambda=8,8,4$$

$\lambda=8$

$$\begin{bmatrix}-3&0&-1\\0&0&0\\-3&0&-1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$$

$-3v_1-v_3=0$

$v_3=-3v_1$

$v_1=e_1$

$v_3=-3e_1$

$v_2=0$

$$\begin{bmatrix}1\\0\\-3\end{bmatrix}\rightarrow e_1$$

$\lambda=4$

$$\begin{bmatrix}1&0&-1\\0&4&0\\-3&0&3\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$$

$v_1-v_3=0$

$v_1=v_3$

$v_1=e_1$

$v_2=0$

$v_3=e_1$

$$\begin{bmatrix}1\\0\\-3\end{bmatrix}\quad\begin{bmatrix}1\\0\\-3\end{bmatrix}\quad\begin{bmatrix}1\\0\\1\end{bmatrix}$$

$$A=\begin{bmatrix}0&1&3\\0&6&0\\-6&2&9\end{bmatrix}$$

$$\begin{bmatrix}-\lambda&1&3\\0&6-\lambda&0\\-6&2&9-\lambda\end{bmatrix}$$

$$-\lambda(6-\lambda)(9-\lambda)-1(0)+3(6(6-\lambda))=0$$
$$-\lambda(6-\lambda)(9-\lambda)+18(6-\lambda)=0$$
$$(6-\lambda)(-9\lambda+\lambda^2+18)=0$$
$$(6-\lambda)(\lambda-6)(\lambda-3)=0$$
$$\lambda=6,6,3$$

$\lambda=6$

$$\begin{bmatrix}-6&1&3\\0&0&0\\-6&2&3\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}$$

$$\begin{bmatrix}-6&1&3\\0&0&0\\0&1&0\end{bmatrix}\quad\begin{bmatrix}-6&0&3\\0&0&0\\0&1&0\end{bmatrix}$$

$v_2=0$

$6v_1=3v_3$

$2v_1=v_3$

$v_1=e_1$

$v_2=0$

$v_3=2e_1$

$$\begin{bmatrix}1\\0\\2\end{bmatrix}$$

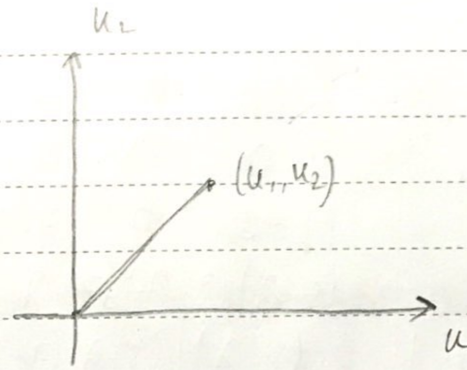
$$\begin{bmatrix}-3&1&3\\0&3&0\\-6&2&6\end{bmatrix}$$

$v_2=0$

$v_1=v_3$

$$\begin{bmatrix}1\\0\\-1\end{bmatrix}$$

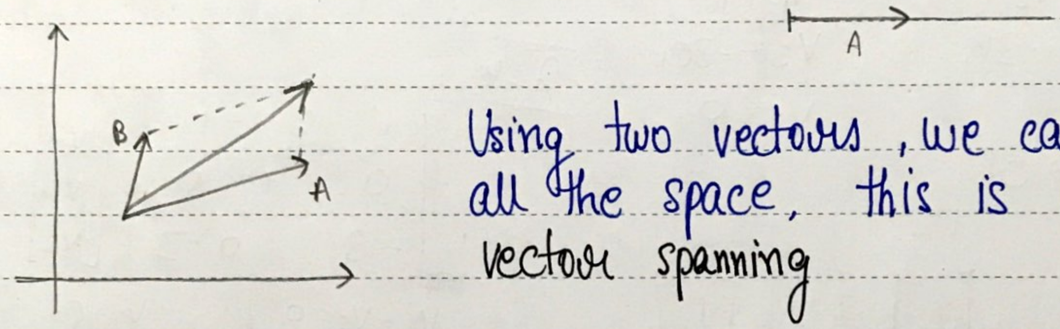
$$\bar{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$



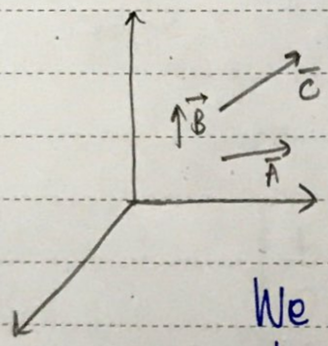
$$u = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix}$$

$$[u_1, u_2 \dots u_n]^T$$

Linearly independent vectors : Not having same direction



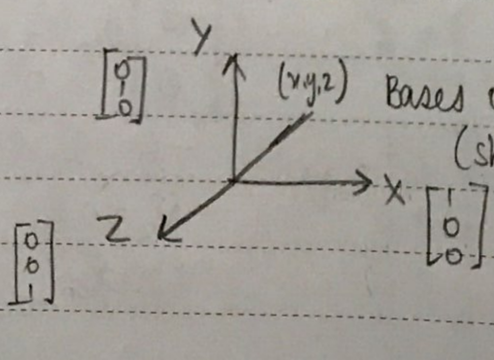
Using two vectors, we can cover all the space, this is called vector spanning



$\vec{C}$ can be a linear combination of $\vec{A}$ and $\vec{B}$. Thus, these vectors are not linearly independent.

We need three linearly independent vector to represent a 3-D space.

For a n-dimension space, we would require n linearly independent vectors to represent the space. And this is called bases.



Bases of a 3-D space
(should be linearly independent)

There can exist numerous bases

Anant Ashutosh Sharma
2K15/CO/026

Nooms

$\|\bar{x}\| > 0$, $\|\bar{x}\| = 0$ if $\bar{x} = \bar{0}$

$\|\alpha \bar{x}\| = |\alpha| \, \|\bar{x}\|$

$\|\bar{x} + \bar{y}\| \leq \|\bar{x}\| + \|\bar{y}\|$

L_2 Norm or Eucledean Norm

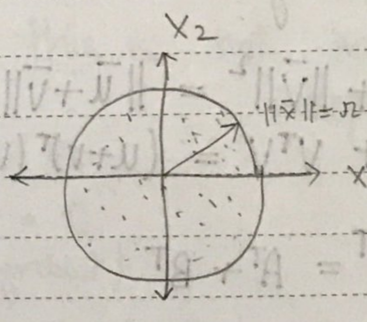
$$\|\bar{x}\|_2 = \left[\sum_{i=1}^{n} |x_i|^2 \right]^{1/2}$$ (Square and then root)

L_1 Norm $\|\bar{x}\|_1 = \left[\sum_{i=1}^{n} |x_i| \right]$ (summation of all coordinates)

L_∞ Norm $\|\bar{x}\|_\infty = \max_{i=1 \dots n} \left[|x_i| \right]$ (Max of all coordinates)

$S_1 = \{ \bar{x} \in \mathbb{R}^2 \mid \|x\|_2 < r \}$

$x_1^2 + y_1^2 = r$

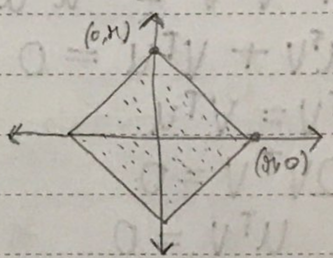


L_p Norm $1 \leq p < \infty$

$$L_p = \left[\sum_{i=1}^{n} |x_i|^p \right]^{1/p}$$

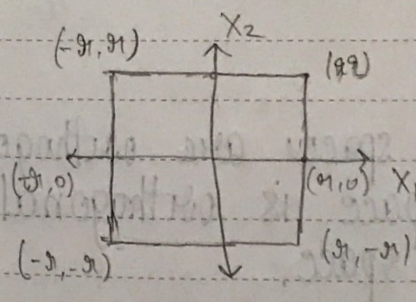
$S_2 = \{ \bar{x} \in \mathbb{R}^2 \mid \|\bar{x}\|_1 < r \}$ → Two dimensional space.

$x_1 + x_2 = r$



$S_3 = \{ \bar{x} \in \mathbb{R}^2 \mid \|x\|_\infty < r \}$

All coordinatao dose to 1



Norm → L_2 Norm (Unless specified)

Inner Product or dot product

$$\bar{A}\cdot\bar{B} = \|\bar{A}\|\,\|\bar{B}\|\cos\theta$$

$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ $A^T B = [a_1 a_2 \cdots a_n]\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3 \cdots a_n b_n$

Thus $\bar{A}\cdot\bar{B} = A^T B$

Also since $\bar{A}\cdot\bar{B} = \bar{B}\cdot\bar{A}$ $\therefore A^T B = B^T A$

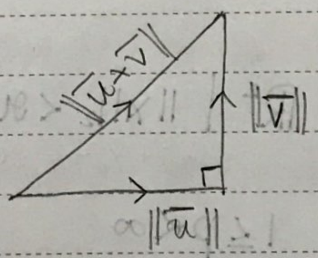
$A^T A = \|A\|^2$

$\|\bar{A}\cdot\bar{B}\| \leq \|\bar{A}\|\cdot\|\bar{B}\|$ Cauchy Swarithz Inequality.

Orthogonality

$\|\bar{u}\|^2 + \|\bar{v}\|^2 = \|\bar{u}+\bar{v}\|^2$

$u^T u + v^T v = (u+v)^T (u+v)$



$[A+B]^T = A^T + B^T$ (Adding corresponding elements)

$\therefore u^T u + v^T v = u^T u + v^T v + u^T v + v^T u$

$u^T v + v^T u = 0$

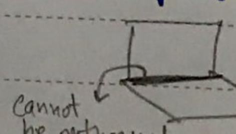
$u^T v = v^T u$

$\therefore 2v^T v = 0$

$\Rightarrow u^T v = 0$ If two vectors are orthogonal, dot products will be zero.

Two sub spaces (S & T) are orthogonal if any (All) vector belonging to S space is orthogonal to any (All) vector belonging to T space.

(Pick any vector)



Subspace can be anything - Point, line, etc
↳ Pass from Origin

For a 2-D plane, a line passing through the origin and perpendicular to the plane will be the only possible orthogonal subspace for the plane. (Complement of Subspace)

Mutual Orthogonality

If we have a set of k vectors $\bar{X}_1, \bar{X}_2 \ldots \bar{X}_k$
They will be mutually orthogonal when

$$\bar{X}_i^T \bar{X}_j = 0 \quad \forall_{i,j} \quad i \neq j$$

Theorem

If they are orthogonal to each other, then they will also be linearly independent.

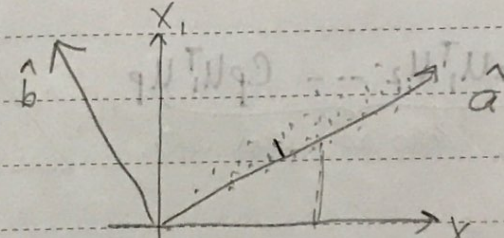
$\sum_{i=1}^{k} \alpha_i \bar{X}_i = 0$ let us assume this is not linearly independent

$$(\alpha_1 \bar{X}_1 + \alpha_2 \bar{X}_2 + \ldots \alpha_k \bar{X}_k)^T \cdot \bar{X}_1 = 0$$

$$\alpha_1 X_1^T X + \underset{0}{\alpha_2 X_2^T X_1} + \underset{0}{\ldots} \underset{0}{\alpha_k X_k^T X_1} = 0$$

$\therefore \quad \alpha_1 \|X_1\|^2 = 0 \quad \alpha_1 = 0$ Since $\|X_i\|^2 \neq 0$

Thus $\alpha_i = 0 \quad \forall i$



change the axis
Reduce dimensionality

$\{\bar{u}_1, \bar{u}_2, \bar{u}_3 \ldots \bar{u}_p\}$

$$\bar{Y} = c_1 \bar{x}_1 + c_2 \bar{u}_2 + \ldots c_n \bar{u}_n$$

$$u^T \bar{Y} = c_1 u_1^T u_1 + \underbrace{c_2 u_1^T u_2 + \ldots c_n u_1^T u_n}_{0}$$

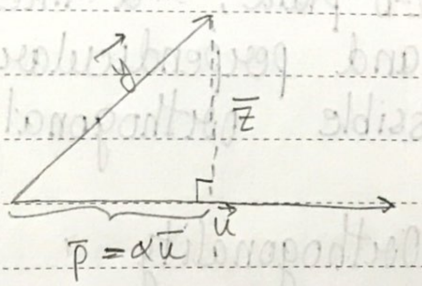
$$c_1 = \frac{u_1^T y}{u_1^T u_1} \qquad c_j = \frac{u_j^T y}{u_j^T u_j}$$

$$\bar{Y} = \frac{u_1^T y}{\|u_1\|^2} \bar{u}_1 + \frac{u_2^T y}{\|u_2\|^2} \bar{u}_2 + \ldots \frac{u_n^T y}{\|u_n\|^2} \bar{u}_n$$

Projection:

$$\bar{p} = \left(\frac{u^T y}{u^T u}\right)\bar{u}$$

$$= \bar{u}\left(\frac{u^T y}{u^T u}\right)$$



$$\bar{p} = \frac{\bar{u}\bar{u}^T}{\bar{u}^T\bar{u}}\bar{y} \qquad \bar{y} = \bar{z} + \alpha\bar{u}$$

$$\bar{z} = \bar{y} - \alpha\bar{u}$$

$\bar{p} = P\bar{y}$

$$u^T(y - \alpha\bar{u}) = 0 \qquad \alpha = \frac{u^T y}{u^T u}$$

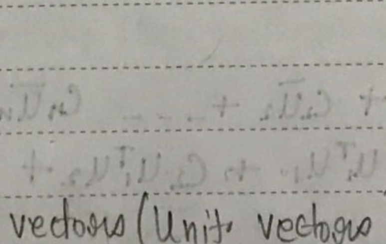
$$\alpha u^T u = u^T y$$

$$u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad V = \begin{bmatrix} 5 \\ 0 \\ 2 \end{bmatrix} \qquad u^T V = V^T u$$

$$uv^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\begin{bmatrix} 5 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 2 \\ 10 & 0 & 4 \\ 15 & 0 & 6 \end{bmatrix}$$

$$\hat{y} = c_1\bar{u}_1 + c_2\bar{u}_2 + \ldots c_p\bar{u}_p$$

(p dimensions) $\{u_1, u_2, u_3 \ldots u_p\}$



$$u_1^T y = c_1 u_1^T u_1 + \underbrace{c_2 u_1^T u_2 \ldots c_p u_1^T u_p}_{\text{Zero as . prod is 0.}}$$

$$\therefore c_1 = \frac{u_1^T y}{u_1^T u_1}$$

$$\therefore c_j = \frac{u_j^T y}{u_j^T u_j}$$

$u_j^T u_j \Rightarrow 1$ for orthogonal vectors (Unit vectors)

$$\hat{y} = \frac{u_1 u_1^T}{u_1^T u_1} y + \frac{u_2 u_2^T}{u_2^T u_2} y + \ldots \frac{u_p u_p^T}{u_p^T u_p} y$$

$$= \bar{u}_1(u_1^T y) + \bar{u}_2(u_2^T y) + \ldots \bar{u}_p(u_p^T y)$$

Linear Combination of basis Vectors.

Syll → 1st Unit
2nd Unit – Normal, Bayse.

Converted into linear combination of coloumn matrix:

$$y = \underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_p \\ | & | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} u_1^T y \\ u_2^T y \\ \vdots \\ u_p^T y \end{bmatrix}}_{U^T y} \rightarrow \begin{bmatrix} — u_1^T — \\ — u_2^T — \\ \vdots \\ — u_p^T — \end{bmatrix} \begin{bmatrix} | \\ y \\ | \end{bmatrix}$$

$$\therefore \hat{y} = UU^T y$$

Change of bases

$B = \{\bar{V}_1, \bar{V}_2 \ldots \bar{V}_k\}$ K-dimension vector space.

$a = \underbrace{c_1\bar{V}_1 + c_2\bar{V}_2}_{\text{Linear Combination}} + \ldots\ldots + c_k\bar{V}_k$ $\quad [a]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$ → Coefficients.

↪ Any n dimensional vector

$\bar{a} \in \mathbb{R}^n$
(n dimensions)

$$[a]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

Vectors.

$$= \underbrace{\begin{bmatrix} | & | & & | \\ V_1 & V_2 & \cdots & V_k \\ | & | & & | \end{bmatrix}}_{C} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

$$\begin{array}{ccc} c_1 & c_2 & c_3 \end{array}$$
$$\begin{bmatrix} 1 & 3 & 7 \\ 4 & 2 & 2 \\ 7 & 4 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix} = 4c_1 + 5c_2 + 8c_3$$

$$\bar{a} = C[a]_B$$

eg: $\bar{V}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\quad \bar{V}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ → 3-dimensional base, but Lies in a plane!!

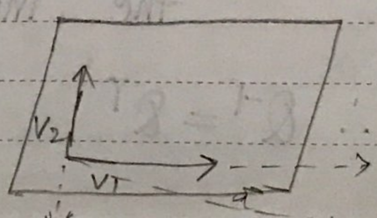
$B = \{\bar{V}_1, \bar{V}_2\}$

$a \in \mathbb{R}^3$ $\quad [a]_B = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

$\vec{a} = \alpha\bar{V}_1 + \beta\bar{V}_2 = 7\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + (-4)\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

$\vec{a} = 7\vec{V}_1 - 4\vec{V}_2$

$$\vec{a} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ -4 \end{bmatrix} = \begin{bmatrix} 3 \\ 14 \\ 17 \end{bmatrix}$$



$\mathbb{R}^3$

$\vec{a} = \begin{bmatrix} 8 \\ -6 \\ 2 \end{bmatrix}$ Vector $\vec{a}$ is also in subspace b (2-dimension)

$B = \{\vec{V}_1, \vec{V}_2\}$ $\quad \vec{V}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\vec{V}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Find coordinate of $\vec{a}$ in the B subspace.

$\vec{a} = B[a]_B \qquad B^{-1}\vec{a} = B^{-1}B[a]_B$ $\qquad \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}$

$[a]_B = B^{-1}\vec{a}$

$\begin{bmatrix} 8 \\ -6 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha+\beta \\ 2\alpha \\ 3\alpha+\beta \end{bmatrix}$ $\quad \alpha = -3$, $\beta = 11$ $\quad \begin{bmatrix} -3 \\ 11 \end{bmatrix}$

Change of basis

$Q = \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} \rightarrow$ These vectors are orthogonal.

$q_i^T \cdot q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

$Q^T = \begin{bmatrix} - & q_1^T & - \\ - & q_2^T & - \\ & \vdots & \\ - & q_n^T & - \end{bmatrix}$ $\qquad QQ^T = \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - q_1^T - \\ - q_2^T - \\ \vdots \\ - q_n^T - \end{bmatrix}$

$$QQ^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

(Orthonormal)

$Q^{-1}Q = I$ Thus, in case of orthogonal fectors, the inverse is same as transpose.

$\therefore Q^{-1} = Q^T$

Anant Ashutosh Sharma
2K15/CO/026

$f(\underset{x}{\uparrow}) \longrightarrow f(x)$

$A\bar{X} = \lambda \bar{X}$ eigen vectors

$A\bar{X} = \bar{b}$ eigen values $\bar{X} \parallel \bar{b}$

↳ linear combination of column vector.

Suppose we have a vector A and n eigen values $(X_1, X_2 \dots X_n)$

$$AS = A \underbrace{\begin{bmatrix} | & | & & | \\ X_1 & X_2 & \dots & X_n \\ | & | & & | \end{bmatrix}}_{S}$$

$AX_1 = \lambda_1 X_1$

$AX_2 = \lambda_2 X_2$

$\vdots$

$AX_n = \lambda_n X_n$

(Not linear combination as it is not being added)

$$\therefore AS = \begin{bmatrix} | & | & & | \\ \lambda_1 X_1 & \lambda_2 X_2 & \dots & \lambda_n X_n \\ | & | & & | \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \bar{X}_1 & \bar{X}_2 & \dots & X_n \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \lambda_n \end{bmatrix}}_{D}$$

D (diagonal matrix) (Made up of eigen values)

$\therefore AS = SD$

(Pre Multiply) $S^{-1}AS = S^{-1}SD$

$\boxed{S^{-1}AS = D}$ To diagonalize matrix A

$ASS^{-1} = SDS^{-1}$

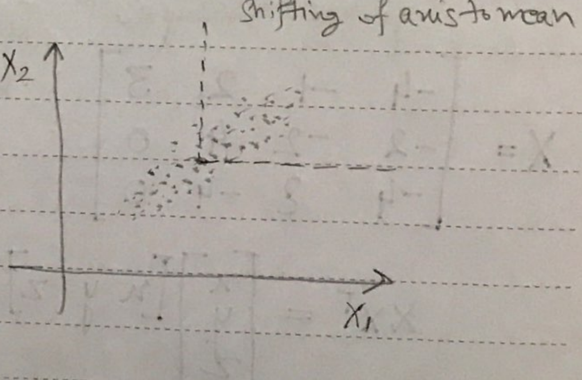
$\therefore \boxed{A = SDS^{-1}}$

* For any symmetric matrix, the eigen values will be real and the eigen vectors will be orthogonal.
and can be made into orthonormal vectors. (Magnitude = 1)

Principle Component Analysis (PCA)

X_1 & X_2 are orthogonal

PCA is used to reduce the dimensionality...



$$\overline{M}_{(\text{Mean})} = \frac{\overline{X}_1 + \overline{X}_2 + \dots \overline{X}_n}{n}$$

$$\text{Variance} = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \mu)^2$$

for $\text{Covar}(XY)$ $\sum (x_i - \mu_x)(y_i - \mu_y)$

↳ for a sample (stating it for population)

After shifting axiss to mean, thus mean is already subtracted from all data

. dot product Square & Add

$$\therefore \text{Variance} = \sum x_i^2 \left(\frac{X^T X}{n-1}\right)$$

$X^T X$

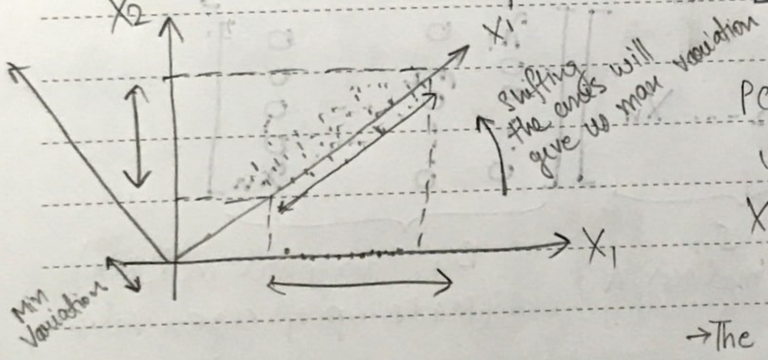
$X^T Y$ Since $\underline{\mu = 0}$ (shifted)

$Y^T X$

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

$$\underset{m \times n}{X} = \begin{array}{c|c} & 1 \dots n \\ \begin{matrix} 1 \\ \vdots \\ m \end{matrix} & \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \end{array}$$

n samples
m features



PCA looks for direction with maximum variance

$X_1' \rightarrow$ has more variation.

(Thus X_1' is a better feature)

→ The more the spread of data, the better.

Earlier, point in standard bases $\overline{a} \in \mathbb{R}^n$

We want to represent this vector in $B = \{\overline{V}_1, \overline{V}_2 \dots \overline{V}_k\}$

$$[\overline{a}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} \qquad \vec{a} = c_1\overline{V}_1 + c_2\overline{V}_2 + \dots c_k\overline{V}_k$$

3 feature, 4 samples

$$X = \begin{bmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{bmatrix}$$

Taking Variance with respect to feature (feature in Rows)

XX^T (because we need to reduce features)

$$XX^T = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} x & y & z \end{bmatrix} = \begin{bmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{bmatrix}$$

Covariance Matrix (Symmetric)

Non diagonal

We want covariance to be 0, we want features to be linearly independent.

By changing bases, we can try to make them 0.

n samples

$$\begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^{n \times p}$$

p features.

$p \times p$ Matrix $\begin{bmatrix} \bar{U}_1 & \bar{U}_2 & \cdots & \bar{U}_p \end{bmatrix}$ Bases Vector.

known, unknown

$$X_1 \quad \underset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}} = \underset{p \times p}{\begin{bmatrix} U_1 & U_2 & \cdots & U_p \end{bmatrix}} \underset{p\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}}$$

1 Person. (New bases) (Space)

$K \to$ points (n)
$m \to$ features (p)
($m \to$ dimental space.)

$X \Rightarrow B$
$Y \Rightarrow G$

$B = UG$

$X = UY$

$Y = U^{-1}X$ So orthonormal vectors

$1 < P$

Symmetric

$$S_X = \frac{1}{n-1} X X^T \quad (\text{Not all variance is } 0)$$

$p \times n$, $p\times p$, $p \times n$

Data Changed: $S_Y = \frac{1}{n-1} YY^T$ (Diagonalize) $= D$ (if we can prove this D, using some bases, we are done)

$$P = \begin{bmatrix} \text{eigen vectors} \\ \text{of } XX^T \end{bmatrix}$$

These will be orthonormal

Changining base in the direction of eigen vectors will lead to covariance 0. !!

if $U = P$

$P^T = P^{-1}$

out of $P \ldots, P-L$ can be reduced

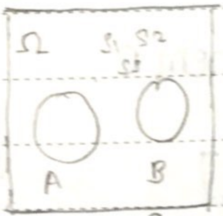
$Y = P^T X$

$$S_Y = \frac{1}{n-1} YY^T = \frac{1}{n-1}\left[P^T X (P^T X)^T\right] = \frac{1}{n-1}\left[P^T X X^T P\right]$$

$$= P^T \frac{XX^T}{n-1} P = P^T S_X P = D$$

(Diagonal Matrix)

Probability Theory



A → Subset of Ω

$P(A) \geq 0$ $\qquad P(A \cup B) = P(A) + P(B)$

$P(\Omega) = 1$

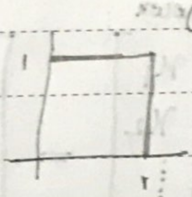
$P(\{s_1, s_2 \ldots s_n\}) = P(\{s_1\}) + P(\{s_2\}) \ldots + P(\{s_n\})$.

$P(A) = \dfrac{\text{no. of elements in } A}{\text{Total no. of points in } \Omega}$

Sum Rule

$P(A_1 \cup A_2 \ldots) = P(A_1) + P(A_2) \ldots$

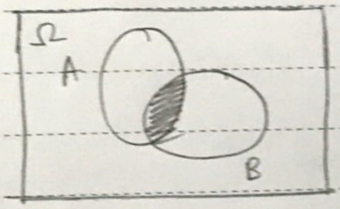
↳ Should be disjoint.



Continuous Sample Space

P(Single Point) Almost 0.

Conditional Probability



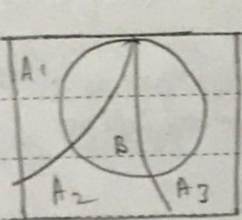
$P(A|B)$ → Same probability, only the sample space is restricted.

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

Product Rule

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

$P(A \cap B \cap C) = P(A).\ P(B/A) \cdot P(C/A \cap B)$



$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

$= P(A_1)\, P(B/A_1) + P(A_2)\, P(B/A_2) + P(A_3)\, P(A/A_3)$

$P(A_i)$

$P(B|A_i)$

$P(A_i/B) = ?$

$P(A_i)\, P(B|A_i) = P(B)\, P(A_i|B)$

$$P(A_i|B) = \frac{P(A_i)\, P(B|A_i)}{P(B)}$$

$$P(A_i|B) = \frac{P(A_i)\, P(B/A_i)}{\sum_j P(A_j)\, P(B|A_j)}$$

Independence of events:

$P(B|A) = P(B)$ When A and B are independent.

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

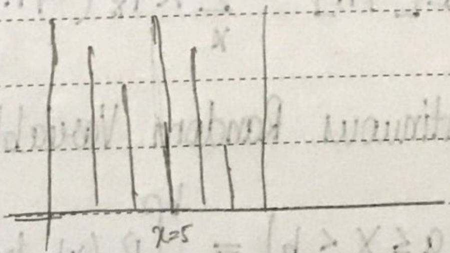
Random Variable

Probability Mass Function (PMF).

$X : \Omega \longrightarrow \mathbb{R}$

$P_X(x) \geq 0$

$\sum_x P_X(x) = 1$

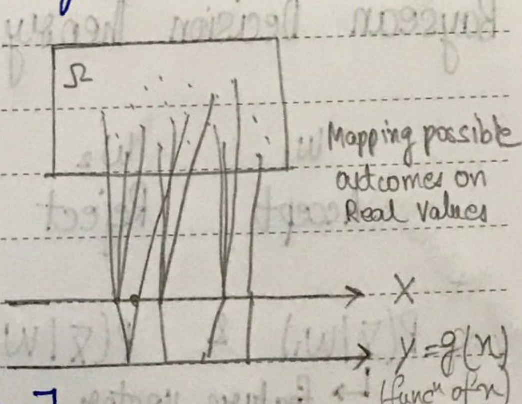


$P_X(x) = P(X = x)$

↳ Value ∈ $\mathbb{R}$

↳ Random variable

PMF: Collect all possible outcomes for which random variable X is equal to x add their probability and repeat it for all x.



Expectation: $E[X] = \sum_x x \, P_X(x)$

$$E[Y] = \sum_y y \, P_Y(y)$$
$$= \sum_x g(x) \, P_X(x) = E[g(x)]$$

In general,
$$E[g(x)] \neq g[E(x)]$$

$E[\alpha] = \alpha$

↳ Single Value (Scalar)

- Variance
- Continuous Variable

$$E[\alpha X] = \alpha E[X]$$

↳ Constant

$$E[\alpha X + \beta] = \alpha E[X] + \beta$$

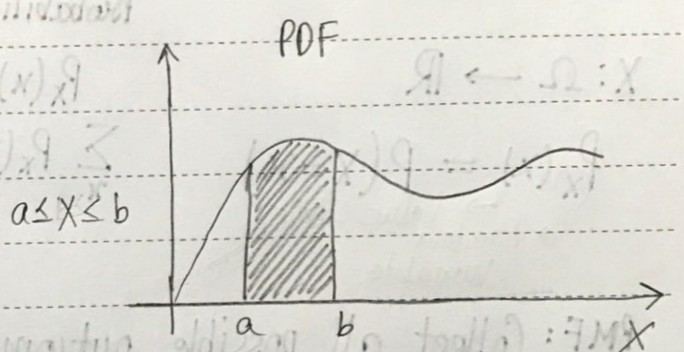
Variance

$$E[X^2] = \sum_x x^2 p_x(x)$$

$$var = \sigma^2 = E[(X - E[X])^2]$$
$$= E[X^2] - (E[X])^2$$

$$E[X/A] = \sum_x x\, P_x(x/A)$$

Continuous Random Variable



$$P(a \le X \le b) = \int_a^b P_x(x)\, dx$$

Baysean Decision Theory

w_1	w_2			Decision
Accept	Reject	if	$P(w_1) > P(w_2) \Rightarrow$	W_1
			$P(w_1) < P(w_2) \Rightarrow$	W_2

(Prior Probability)

$P(\bar{x}/W_1)$ & $P(\bar{x}/W_2)$ → Probability of feature vector in given class

↳ feature vector

$P(w_1/\bar{x})$ & $P(w_2/\bar{x})$ → Probability of class for the given feature vector

Posterior Probability

if $P(w_1/\bar{x}) > P(W_2/\bar{x}) \Rightarrow W_1$

$P(w_1|\bar{x}) < P(W_2|\bar{x}) \Rightarrow W_2$

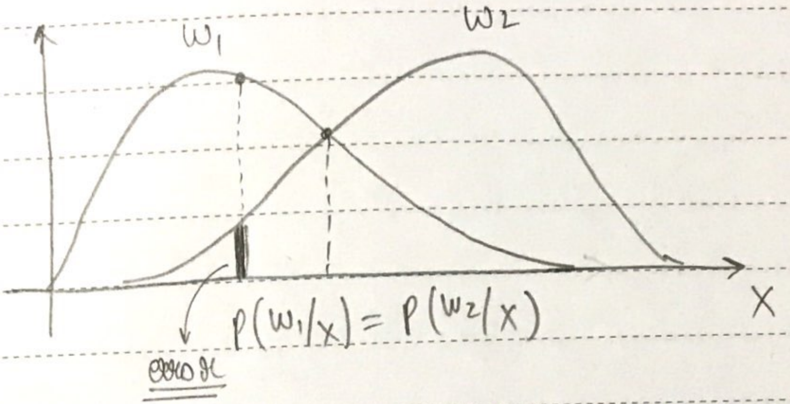
Bayes Rule will be applied here

$$p(w_1)\, p(x|w_1) > p(w_2)\, p(x/w_2)$$

if $P(x/w_1) = P(x/w_2)$

Same feature probable in both

The it is a useless feature :,(



Decision Boundary will be when/at $P(\omega_1/X) = P(\omega_2|X)$

$$P(\text{error}|x) = \begin{cases} P(\omega_2|x) & \text{if } \omega_1 \text{ is selected} \\ P(\omega_1|x) & \text{if } \omega_2 \text{ is selected} \end{cases}$$

$$= \min\{P(\omega_1/x), P(\omega_2/x)\}$$

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}, x)\,dx = \int_{-\infty}^{\infty} P(\text{error}|x)\,P(x)\,dx$$