Applied Machine Learning

Dr. Harikrishnan N B
Computer Science and Information Systems

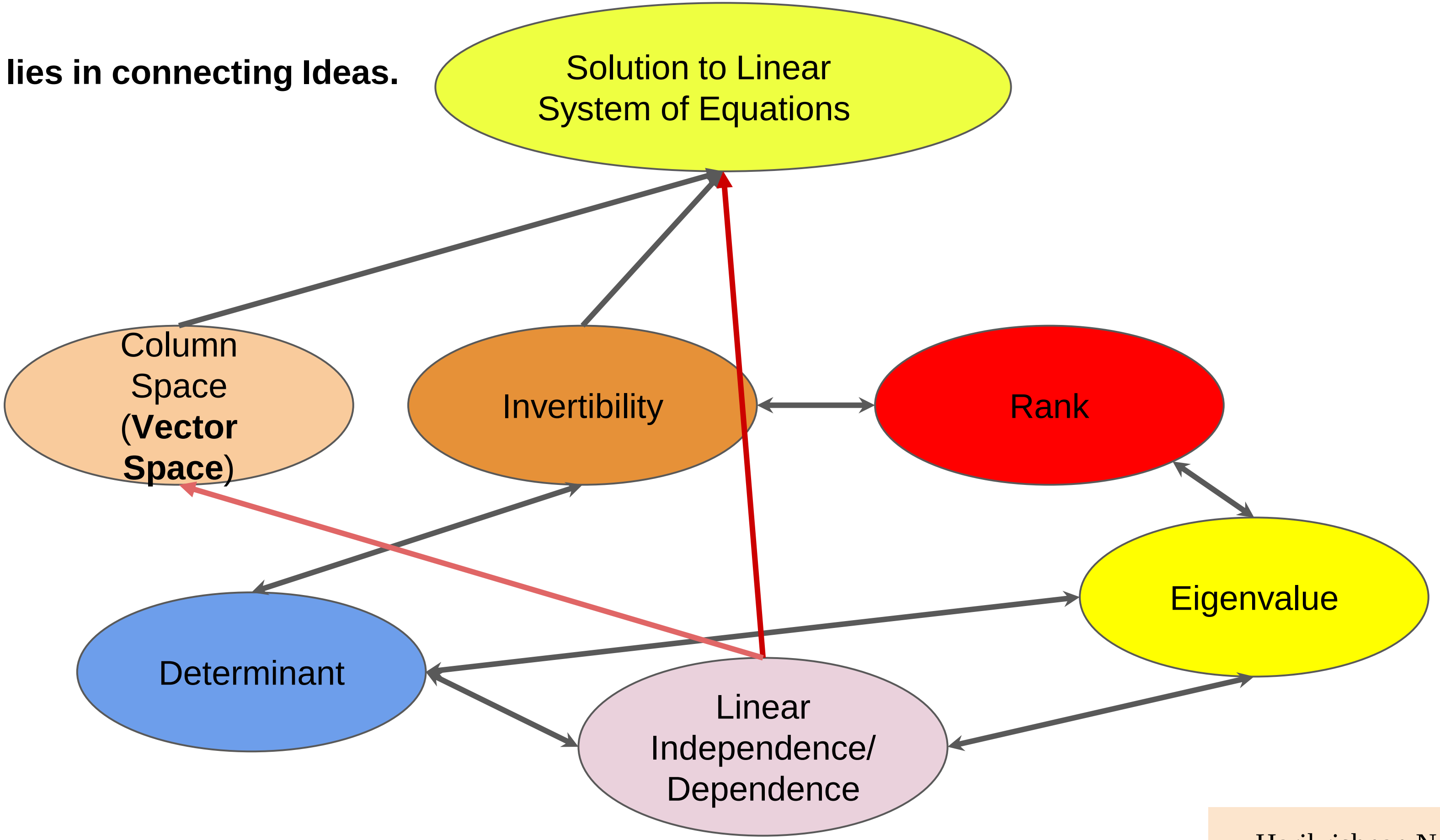
BITS Pilani
Pilani Campus

SE ZG568 / SS ZG568,
Applied Machine Learning
Lecture No. 6 [23- Feb-2025]

Recap

Basics of Linear Algebra,
Row Picture, Col Picture, Algebraic Way
Solution to System of Linear Equations
Inverse of a Matrix
Linear Regression,
PCA

Beauty lies in connecting Ideas.



Practical Challenges and Important Points

When can we apply $A = X\Lambda X^{-1}$ **?**

- A should be a square matrix
- When A has 'n' linearly independent eigenvectors, then X^{-1} always exist.

What happens when A is Symmetric ($A^T = A$)?

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal.
- For an ORTHONORMAL matrix X, the inverse is its transpose $X^{-1} = X^T$

$$A = X\Lambda X^{-1}$$
$$A = X\Lambda X^T$$

Harikrishnan N B

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal. Why?

Practical Challenges and Important Points

When can we apply $A = X\Lambda X^{-1}$?

- A should be a square matrix
- When A has ‘n’ linearly independent eigenvectors, then X^{-1} always exist.

What happens when A is Symmetric ($A^T = A$)?

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal.
- For an ORTHONORMAL matrix X, the inverse is its transpose $X^{-1} = X^T$

$$A = X\Lambda X^{-1}$$
$$A = X\Lambda X^{T}$$

Harikrishnan N B

Practical Challenges

What if A is not a square matrix?

- We cannot apply Spectral Decomposition.

Don’t Worry!!!

Singular Value Decomposition works for any Matrix.

A Few more steps to PCA

What all minimum can we say about this data?

Data instance 1   Data instance 2

X	1	2	3	4	5
Y	1	5	4	6	7

Height $f_1 \rightarrow$ X

Weight $f_2 \rightarrow$ Y

X, Y are the features

What all minimum can we say about this data?

Mean(X) = $$\frac{1}{N}\sum_{i=1}^{N} x_i$$

Variance(X) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$$

Cov(X,Y) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$$

X	Y	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
1	1				
2	5				
3	4				
4	6				
5	7				
Mean (X)	Mean(Y)	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
?	?	?	?	?	?

Harikrishnan N B

Mean(X) = $\frac{1}{N}\sum_{i=1}^{N} x_i$

Variance(X) = $\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$

Cov(X,Y) = $\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$

$\mu_x = \text{Mean}(X) = \frac{1+2+3+4+5}{5} = 3$

$\mu_y = \text{Mean}(Y) = \frac{1+5+4+6+7}{5} = \frac{23}{5} = 4.6$

$\text{Var}(X) = 2.5$, $\text{Var}(Y) = 5.3$

X	Y
1	1
2	5
3	4
4	6
5	7

Var(X)	Var(Y)	
$(1-3)^2 = 4$	$(1-4.6)^2$	$= 12.96$
$(2-3)^2 = 1$	$(5-4.6)^2$	$= 0.16$
$(3-3)^2 = 0$	$(4-4.6)^2$	$= 0.36$
$(4-3)^2 = 1$	$(6-4.6)^2$	$= 1.96$
$(5-3)^2 = 4$	$(7-4.6)^2$	$= 5.76$
$\frac{10}{4}$		

cov(X,Y)

$\cancel{4 \times 12.96}$ $(1-3) \times (1-4.6)$

$(2-3) \times (5-4.6)$

$(3-3) \times (4-4.6)$

$(4-3) \times (6-4.6)$

$(5-3) \times (7-4.6)$

$= 3.25$

cov(Y,X) $= 3.25$

Harikrishnan N B

Mean(X) = $$\frac{1}{N}\sum_{i=1}^{N} x_i$$

Variance(X) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$$

Cov(X,Y) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$$

X	Y
1	1
2	5
3	4
4	6
5	7

What all minimum can we say about this data?

Mean(X) = $$\frac{1}{N}\sum_{i=1}^{N} x_i$$

Variance(X) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$$

Cov(X,Y) = $$\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$$

X	Y	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
1	1				
2	5				
3	4				
4	6				
5	7				
Mean (X)	Mean(Y)	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
3.0	**4.6**	**2.5**	**5.3**	**2.25**	**2.25**

Harikrishnan N B

$$D = \text{Data} = \begin{bmatrix} X & 2 & 4 & \cdots \\ Y & 3 & 5 & \end{bmatrix}$$

$$D_{mean} = \begin{bmatrix} X-\mu_x & 2-\mu_x & 4-\mu_x & \cdots \\ Y-\mu_y & 3-\mu_y & 5-\mu_y & \cdots \end{bmatrix}_{2\times N}$$

$$\frac{1}{N-1}\left(D_{mean} D_{mean}^T\right)_{2\times 2}$$

$$\frac{1}{N-1}\left(D_{mean} D_{mean}^T\right) \Rightarrow \begin{bmatrix} (x_1-\mu_x) & x_2-\mu_x & \cdots & x_n-\mu_x \\ (y_1-\mu_y) & (y_2-\mu_y) & \cdots & (y_n-\mu_y) \end{bmatrix}_{2\times N} \begin{bmatrix} x_1-\mu_x & y_1-\mu_y \\ x_2-\mu_x & y_2-\mu_y \\ \vdots & \vdots \\ x_n-\mu_x & y_n-\mu_y \end{bmatrix}_{N\times 2}$$

$$\frac{1}{N-1}\begin{bmatrix} \sum_{i=1}^{N}(x_i-\mu_x)^2 & \\ & \end{bmatrix}_{2\times 2} \quad \sum_{i=1}^{N}$$

$$\frac{1}{N-1}\begin{bmatrix} \operatorname{Var}(X) & \boxed{\operatorname{Cov}(X,Y)} \\ \operatorname{Cov}(Y,X) & \operatorname{Var}(Y) \end{bmatrix} \rightarrow \sum_{i=1}^{N}\left(x_i-\mu_x\right)\left(y_i-\mu_y\right)$$

Variance- Covariance Matrix

$[\quad]_{2\times 2}$

$A = U\Lambda U^T$

$\frac{1}{N-1} D_{mean} D_{mean}^T$

$$
\begin{array}{ccc}
 & X & Y \\
\begin{array}{c} X \\ Y \end{array} & \multicolumn{2}{c}{\begin{bmatrix} var(X) & cov(X,Y) \\ cov(Y,X) & var(Y) \end{bmatrix}}
\end{array}
$$

A =

Recall the properties of a symmetric matrix!!!

$$
\begin{array}{cc}
 & \begin{array}{ccc} X & Y & Z \end{array} \\
\begin{array}{c} X \\ Y \\ Z \end{array} & \begin{bmatrix} Var(X) & cov(X,Y) & cov(X,Z) \\ cov(Y,X) & var(Y) & cov(Y,Z) \\ cov(Z,X) & cov(Z,Y) & var(Z) \end{bmatrix}
\end{array}
$$

- **Variance - Covariance Matrix is symmetric. cov(X,Y) = cov(Y,X)**
- **The diagonal entries represents variance**
- **The off- diagonal entries represents the correlation of X and Y**

What does Variance - Covariance Matrix signifies?

Case I

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

Case II

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

Case III

$$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

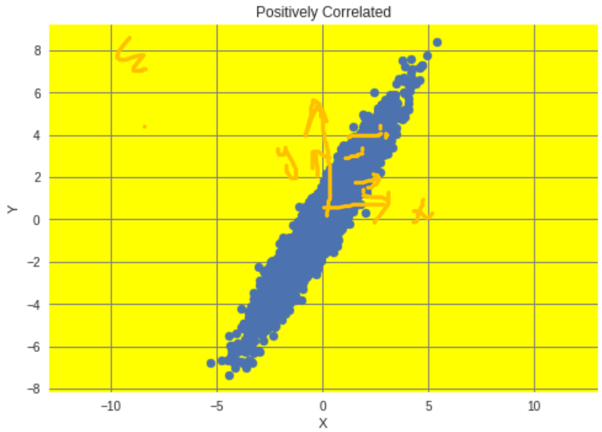
Note: In all cases mean is (0,0)

What does Variance - Covariance Matrix signifies?

$p(x,y) = p(x)\,p(x)$

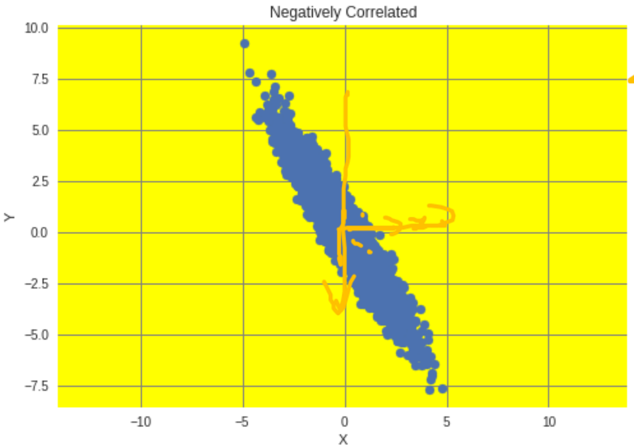
Case I

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



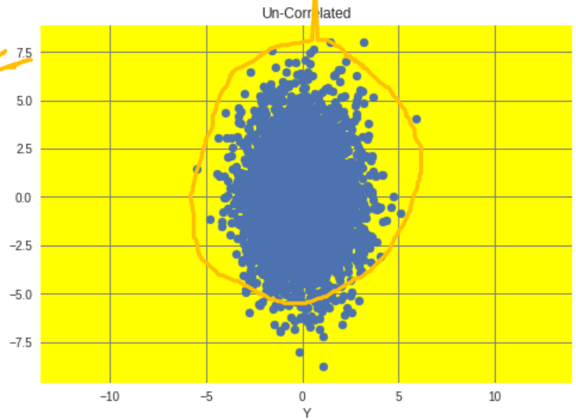
Case II

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



Case III

$$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

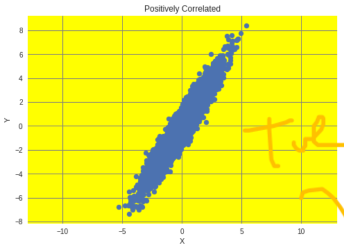


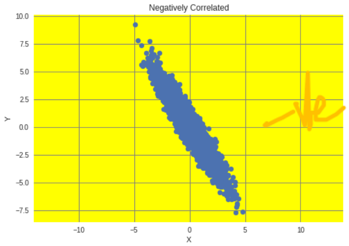
Note: In all cases mean is (0,0)

Harikrishnan N B

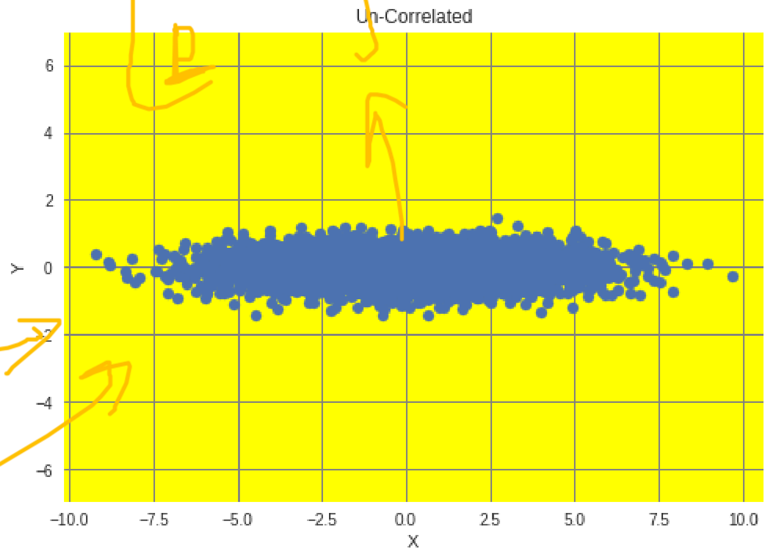
So what does PCA do ?

- Principal Component Analysis (PCA) makes the data **UNCORRELATED.**





PCA achieves this by
Change of Basis

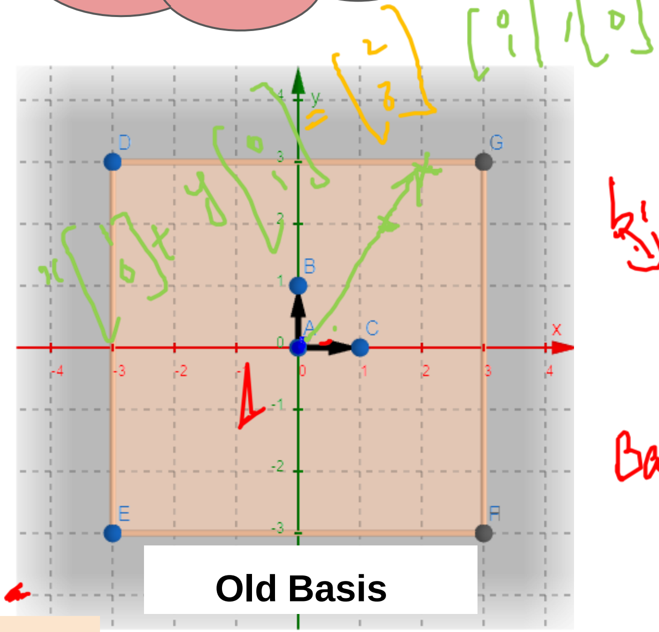


Harikrishnan N B

Change of Basis

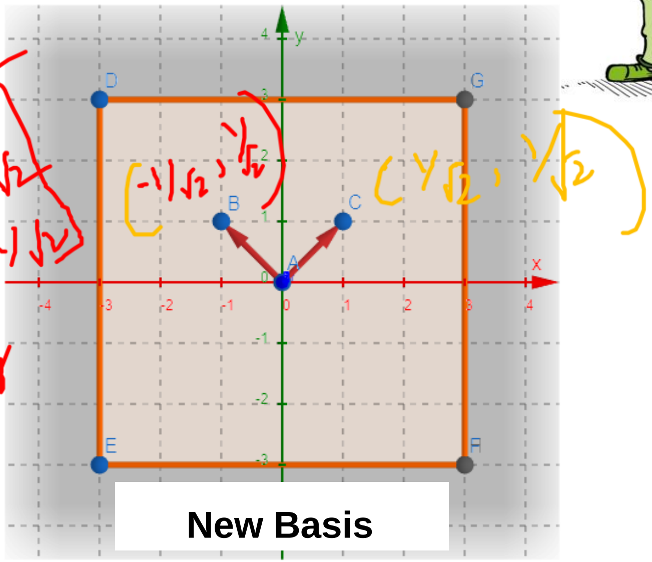
My life is full of struggle.

IT depends on the way you look at life. Try to see life from a different **POINT OF VIEW**



Old Basis

Basis Vector



New Basis

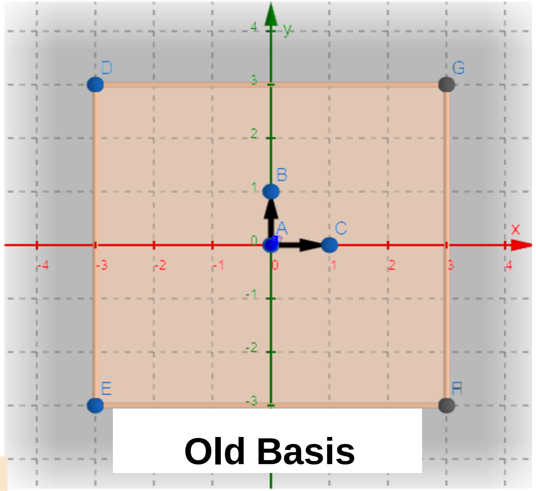
Harikrishnan N B

Recall

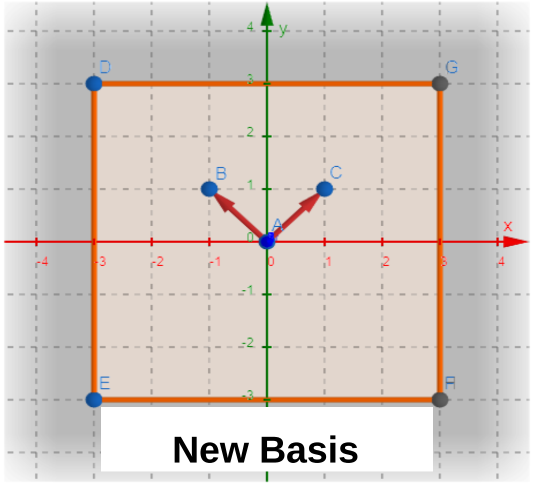
Dimension of a Vector space - Every vector space has a dimension. Dimension is the number of basis vectors required to span the vector space.

Properties of Basis Vectors -

- Basis vectors has to be linearly independent.
- Basis vectors should span the vector space.



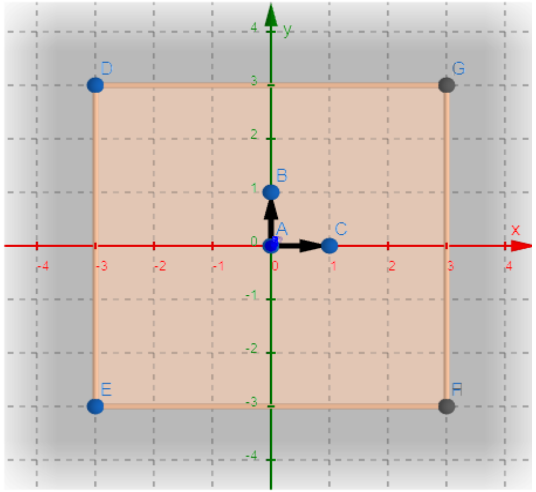
Old Basis



New Basis

Example of Change of Basis

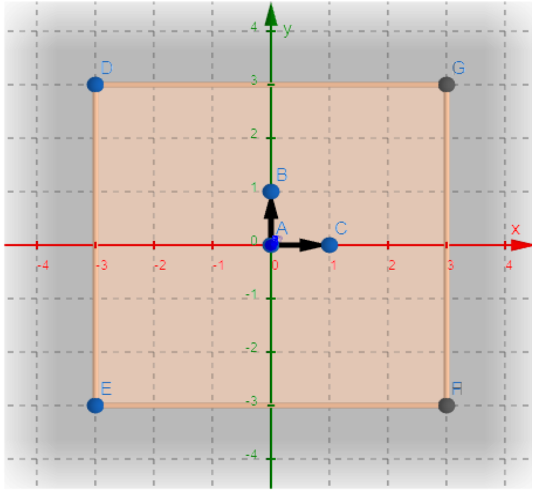
To represent a point (2,3) in old basis and new basis- How to understand this?



Old basis = $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

Example of Change of Basis

To represent a point (2,3) in old basis and new basis- How to understand this?



Old basis = $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

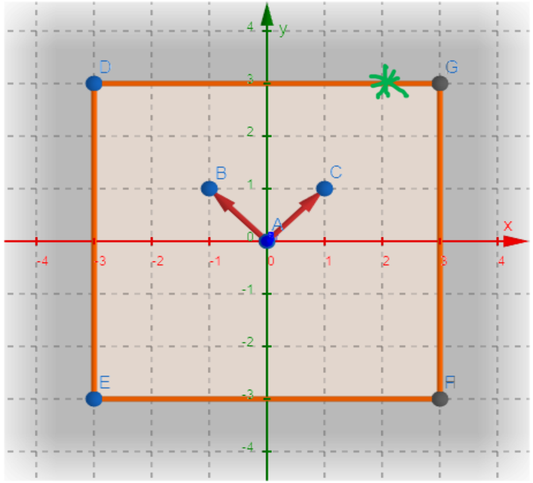
$$2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$b_1 \quad b_2$

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

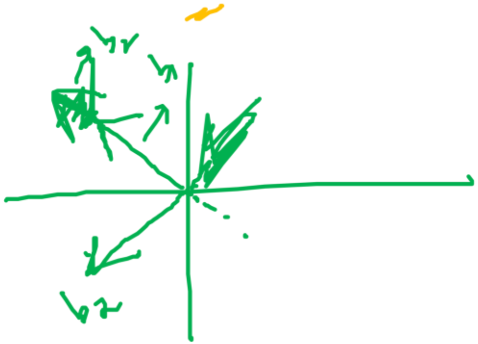
$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

New Basis Representation



New Basis $= \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$

$$x \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + y \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



Harikrishnan N B

$$b_1 = \begin{bmatrix}1\\1\end{bmatrix} \qquad b_2 = \begin{bmatrix}1\\-1\end{bmatrix}$$

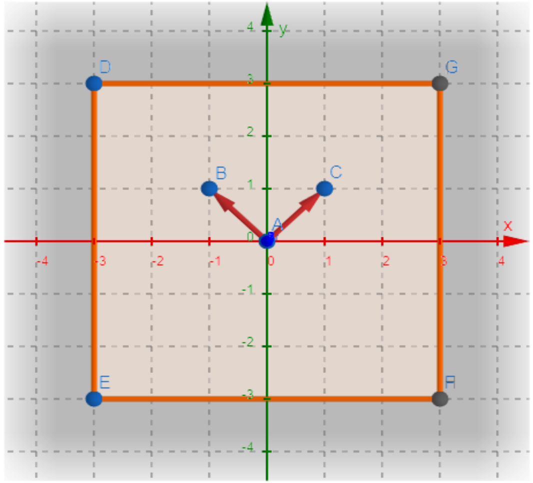
$$\begin{bmatrix}1&1\\1&-1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}a\\b\end{bmatrix}$$

$$\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}1&1\\1&-1\end{bmatrix}^{-1}\begin{bmatrix}a\\b\end{bmatrix}$$

$$\begin{bmatrix}1&1\\1&-1\end{bmatrix}\cdot\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

$$x\begin{bmatrix}1\\1\end{bmatrix} + y\begin{bmatrix}1\\-1\end{bmatrix} = \begin{pmatrix}2\\0\end{pmatrix}$$

Finding x and y for representing (2,3) using new basis

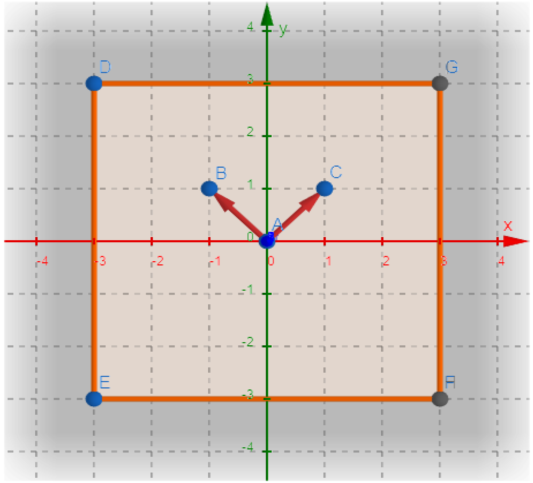


$$x\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} + y\begin{bmatrix}\frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}2\\ 3\end{bmatrix}$$

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 3\end{bmatrix}$$

$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 3\end{bmatrix}$$

Harikrishnan N B



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

$$\underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{P} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$P\vec{x} = \vec{y}$$

$$P^{-1}P\vec{x} = P^{-1}\vec{y}$$

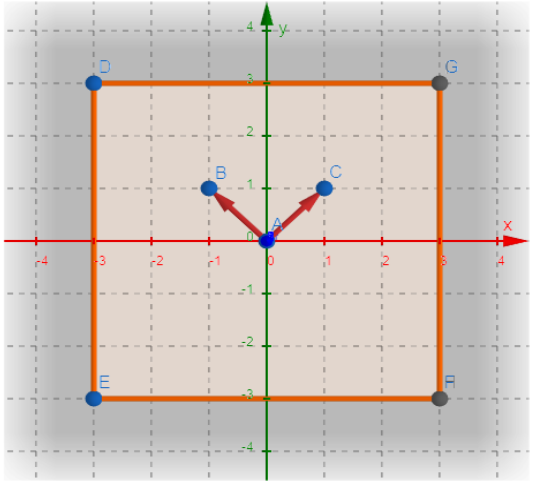
$$\vec{x} = P^{-1}\vec{y}$$

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

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

For **ORTHONORMAL MATRIX, $P^{-1} = P^{T}$**

In our case the matrix P is ORTHONORMAL



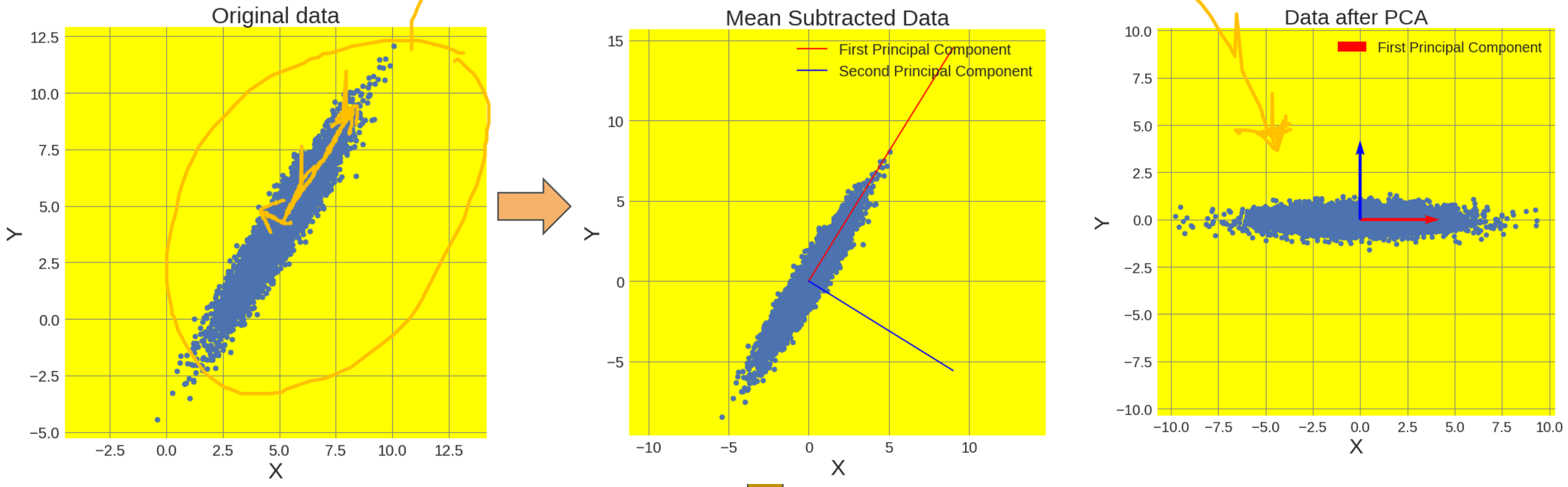
$$\vec{x} = P^{-1}\vec{y} = P^T\vec{y}$$

$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\boxed{\frac{5}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}}$$

Harikrishnan N B

Steps in PCA



Change of Basis

$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{bmatrix} \vec{x_1} & \vec{x_2} & \vec{x_3} & \cdots & \vec{x_n} \end{bmatrix} = \begin{bmatrix} \vec{y_1} & \vec{y_2} & \vec{y_3} & \cdots & \vec{y_n} \end{bmatrix}$$

Harikrishnan N B

$$D_{mean} = \begin{bmatrix} x_1 - \mu_x & x_2 - \mu_x & \cdots & x_n - \mu_x \\ y_1 - \mu_y & y_2 - \mu_y & & y_n - \mu_y \end{bmatrix}$$

$$PD_{mean} = Z \rightarrow$$ linear uncorrelated

Linearly Uncorrelated

$$P_{2\times 2} D_{mean} = Z$$

$$cov(Z) = \frac{1}{N-1}$$

- var-cov matrix of Z

$$(AB)^T = B^T A^T$$

$$\frac{1}{N-1} Z^T_{2\times N} Z_{N\times 2}$$

$$= \frac{1}{N-1} PD_{mean} (PD_{mean})^T$$

$$= \frac{1}{N-1} PD_{mean} D^T_{mean} P^T$$

$$= \frac{1}{N-1} P D_m D_m^T P^T$$

$$\begin{array}{c|cc} & x & y \\ \hline x & Var(x) & cov(x,y) \\ y & cov(y,x) & Var(y) \end{array}$$

$$P \underbrace{\left[\frac{1}{N-1} D_m D_m^T\right]}_{} P^T \rightarrow \text{symmetric matrix}$$

$$A = V \Lambda V^T$$

$$\boxed{\begin{array}{l} V^T V = I \\ V V^T = I \end{array}}$$

$$Var\text{-}cov(Z) = P V \Lambda V^T P^T$$

V is the eigenvector variance covariance matrix of Data

V is orthonormal.

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

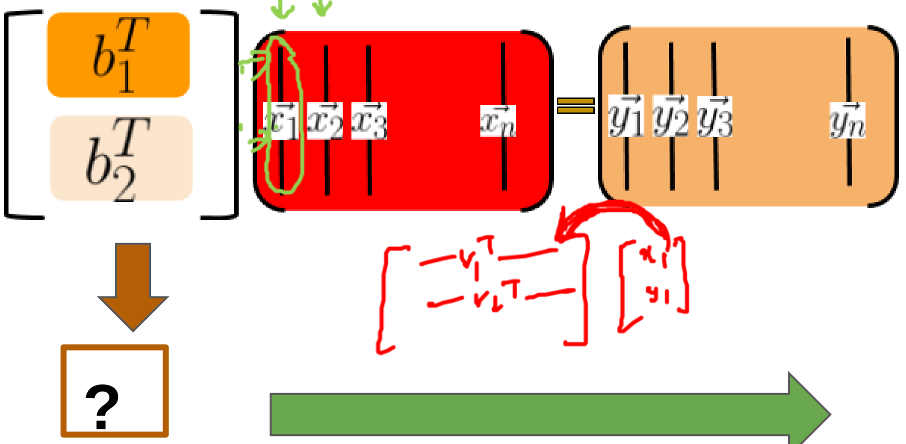
$$I V^T V \Lambda V^T V I$$

P is unknown

What shall be my P so that the anti-diagonal entries of var-cov(Z) is 0.

What should be the NEW BASIS so that DATA is UNCORRELATED?

D_{mean} Z



$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$

Rows of matrix P are the **eigenvectors** of the **variance - covariance matrix** of the **mean subtracted data**

$$\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ & \vdots & \end{bmatrix}$$

$$PX = Y \qquad D\, D_{mean} = Z$$

$$\text{var-}cov(Y) = cov(PX)$$

$$\text{var-}cov(PX) = \frac{1}{N-1}(PX)(PX)^T$$

$$\text{var-}cov(PX) = \frac{1}{N-1}PXX^TP^T$$

$$\text{var-}cov(PX) = P(\frac{1}{N-1}XX^T)P^T$$

$$\text{var-}cov(PX) = P\,\text{var-}cov(X)P^T$$

$$\text{var-}cov(PX) = P(V\Lambda V^T)P^T$$

$$P = V^T$$

$$\text{var-}cov(PX) = \Lambda$$

Some words about PCA

- PCA is "an orthogonal linear transformation that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (*first principal component*), the second greatest variance lies on the second coordinate (*second principal component*), and so on."

Applications of PCA

number of features = 10

- Dimensionality Reduction
- Denoising
- Feature Extraction
- Image Compression
- EEG Analysis

$$P_{10\times 10} = \begin{bmatrix} - v_1^T - \\ - v_2^T - \\ \vdots \\ - v_{10}^T - \end{bmatrix}$$

$P_{10\times 10}\, D_{mean\; 10\times 100} = Z_{10\times 100}$

Data instance

$Z[0:2, :]$ → first two rows of Z

first row of Z ⇒ first PC

second row of Z ⇒ second PC

Harikrishnan N B

Assumptions in PCA

$$\hat{f}_1 = \frac{f_1 - \min(f_1)}{\max(f_1) - \min(f_1)}$$

$$\hat{f}_2 = \frac{f_2 - \min(f_2)}{\max(f_2) - \min(f_2)}$$

compact size of cell

$$\begin{array}{cc} f_1 & f_2 \\ \end{array}$$
$$\begin{bmatrix} \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \end{bmatrix}$$

- Linearity
- Large variance have important structure
- Principal components are orthogonal

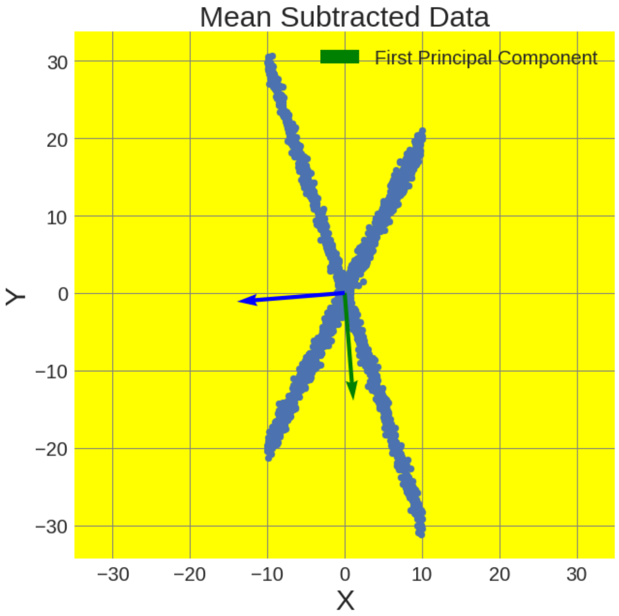
$$\hat{f}_1 \in [0, 1]$$

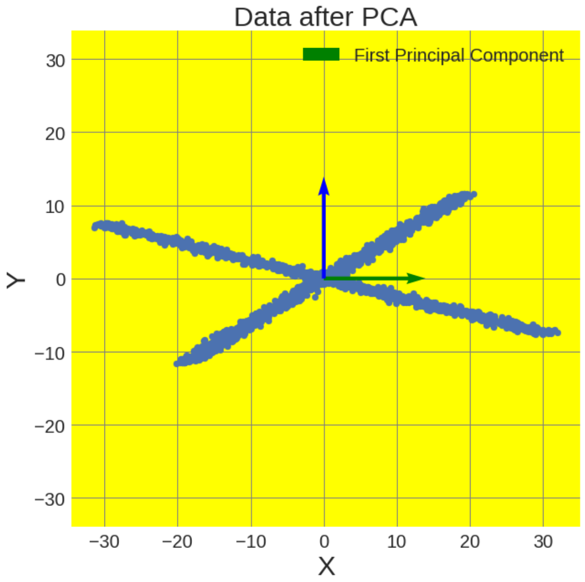
$$\hat{f}_2 \in [0, 1]$$

Harikrishnan N B

When does PCA fail?

- Non-linearity
- Non-Gaussian
- Non-orthogonality

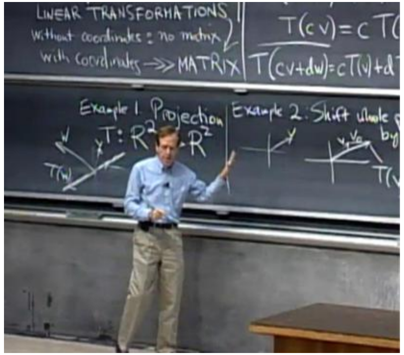


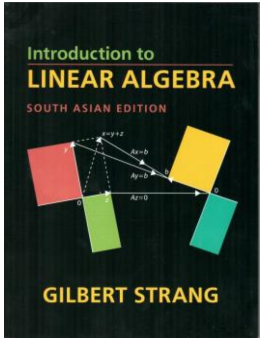


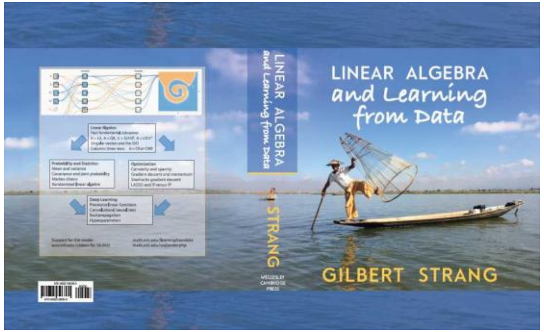
Ref: https://arxiv.org/abs/1404.1100

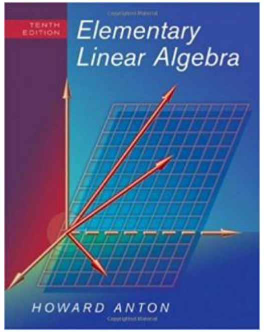
Interesting Materials

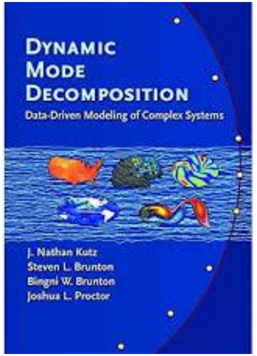
Prof. Gilbert Strang











Tutorial on PCA - [(Click here)]

Learning - Function Approximation

Problem Setting

- Set of possible instances *X*.
- Unknown target function *f: X—> Y*
- Set of function hypotheses *H* = *{h| h: X—> Y}*

Input

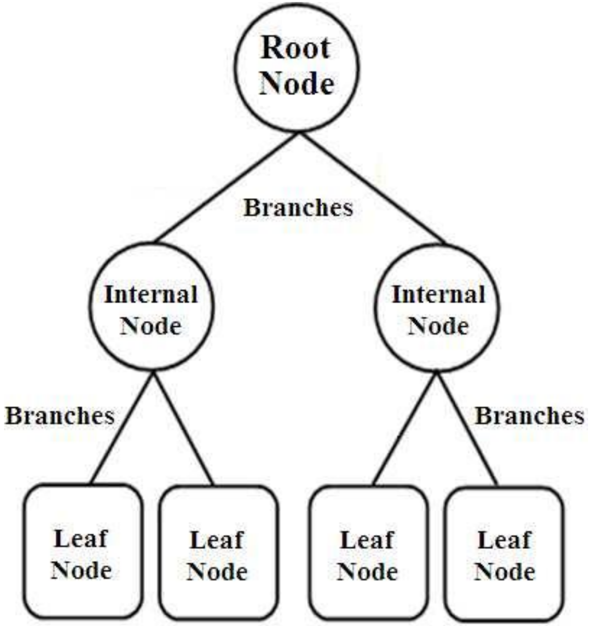
Training Examples of unknown target function *f.*

Output

Hypothesis $h \in H$ that best approximates target function *f.*

Decision Tree

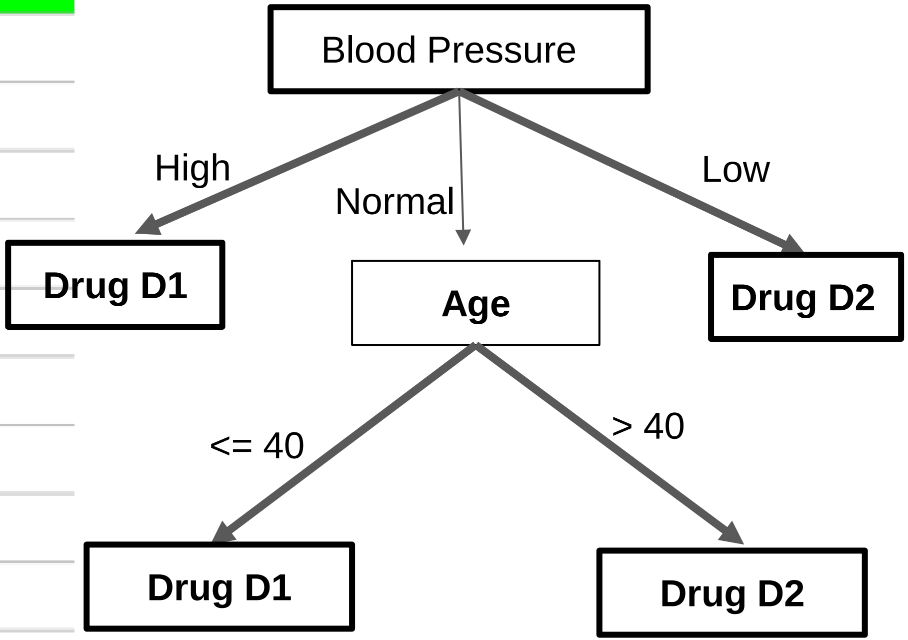
Decision Tree Learning is a method for approximating the target function **(*Y*)**, in which the **learned functions (h)** is represented by a **decision tree**.



h*: *X*—> *Y

ID No	SEX	AGE	Blood Pressure	Drug
1	M	20	Normal	D1
2	F	73	Normal	D2
3	M	37	High	D1
4	M	33	Low	D2
5	F	48	High	D1
6	M	29	Normal	D1
7	F	52	Normal	D2
8	M	42	Low	D2
9	M	61	Normal	D2
10	F	30	Normal	D1
11	F	26	Low	D2
12	M	54	High	D1

ID No	SEX	AGE	Blood Pressure	Drug
1	M	20	Normal	D1
2	F	73	Normal	D2
3	M	37	High	D1
4	M	33	Low	D2
5	F	48	High	D1
6	M	29	Normal	D1
7	F	52	Normal	D2
8	M	42	Low	D2
9	M	61	Normal	D2
10	F	30	Normal	D1
11	F	26	Low	D2
12	M	54	High	D1



Decision Tree

- **ID3** (Iterative Dichotomiser 3)- **1986 Ross Quinlan**
- **C4.5** - is the successor to ID3 and removed the restriction that features must be categorical by dynamically defining a discrete attribute (based on numerical variables) that partitions the continuous attribute value into a discrete set of intervals.
- **C5.0** - is Quinlan's latest version release under a proprietary license. It uses less memory and builds smaller rulesets than C4.5 while being more accurate.
- **CART** - Classification and Regression Trees - s very similar to C4.5, but it differs in that it supports numerical target variables (regression) and does not compute rule sets. CART constructs binary trees using the feature and threshold that yield the largest information gain at each node.

scikit-learn uses an optimized version of the CART algorithm; however, the scikit-learn implementation does not support categorical variables for now.

https://scikit-learn.org/stable/modules/tree.html#classification

ID3- Decision Tree algorithm for classification

How do we construct such a tree?

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No