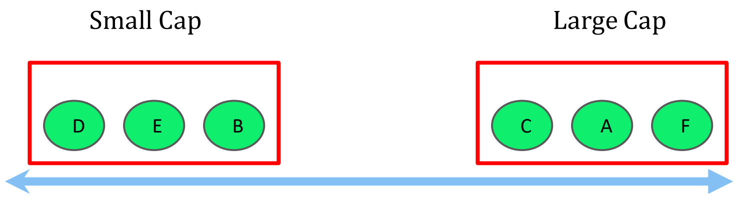
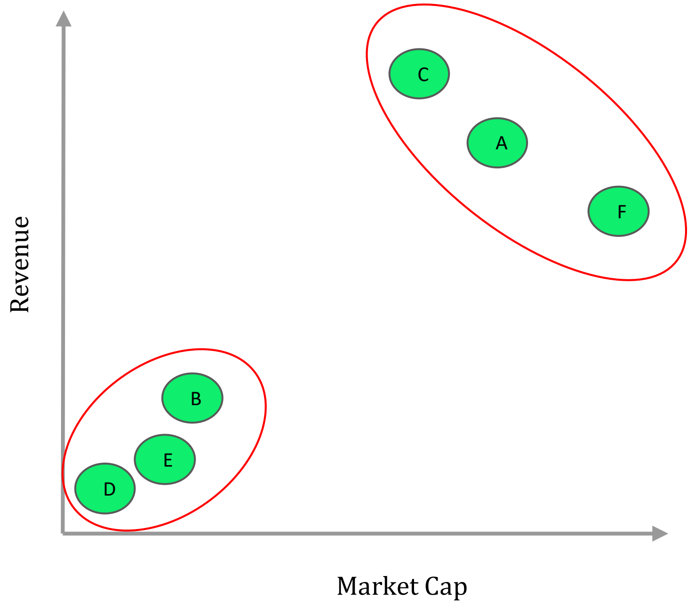
Principal Component Analysis (PCA)

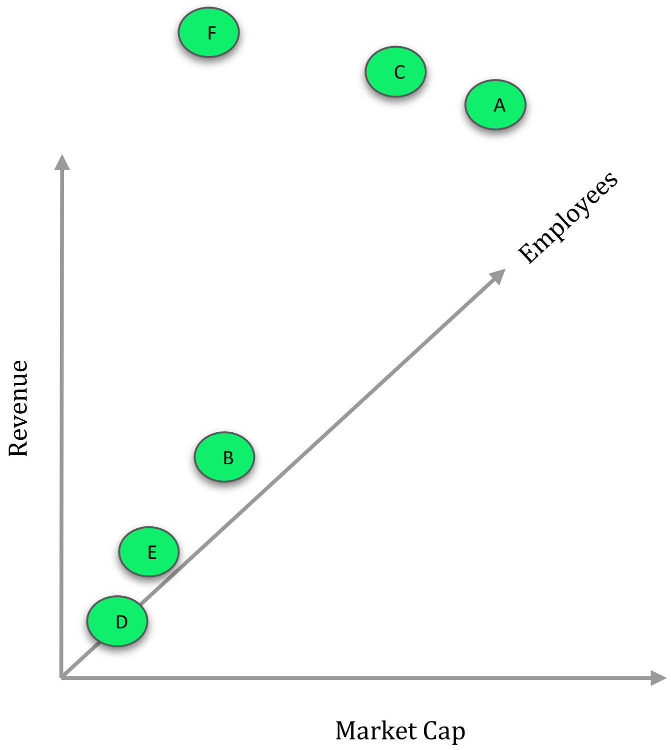
Company	Market Cap ($M)
A	3000
B	120
C	2500
D	10
E	72
F	4000



Company	M-Cap ($M)	Revenue ($K)
A	3000	400
B	120	30
C	2500	420
D	10	3
E	72	20
F	4000	390



Compan y	M-Cap ($M)	Revenue ($K)	Employees
A	3000	400	42000
B	120	30	200
C	2500	420	13000
D	10	3	32
E	72	20	120
F	4000	390	1900



Company	M-Cap ($M)	Revenue ($K)	Employees	Number of Countries
A	3000	400	42000	32
B	120	30	200	7
C	2500	420	13000	59
D	10	3	32	4
E	72	20	120	1
F	4000	390	1900	115

?

Height	Alturo
6	595
4.5	460
7	700
3	290
10	995
12	1210

How many Dimensions?

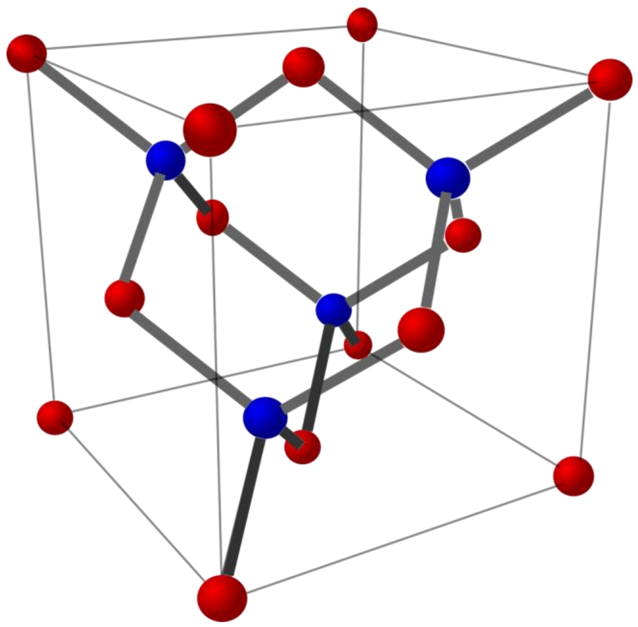
Skidding Accidents	Snow Plow Expense	School Closures	Water Pipe Bursts	Patients with Heat Stroke	Geographic Area
.	.	.	..	...	A
...	...	..	...	..	B
...	..	..	..	.	C
.	.	.	.	...	A

How many Dimensions?

Skidding Accidents	Snow Plow Expense	School Closures	Water Pipe Bursts	Patients with Heat Stroke	Geographic Area
.	.	.	..	...	A
...	...	..	...	..	B
...	..	..	..	.	C
.	.	.	.	...	A

Temperature?

True Dimensionality << Observed Dimensionality



Curse of Dimensionality

Vision

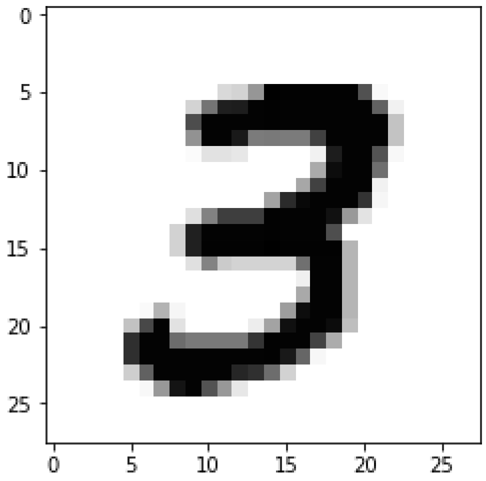
1000 * 1000 Pixels

Text Documents

11 Billion Phrases in English

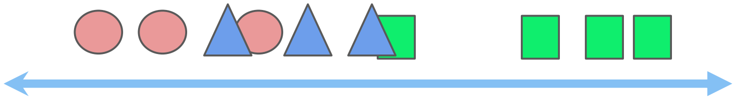
How many Dimensions?

How many possible values?



What possible problem does it create?

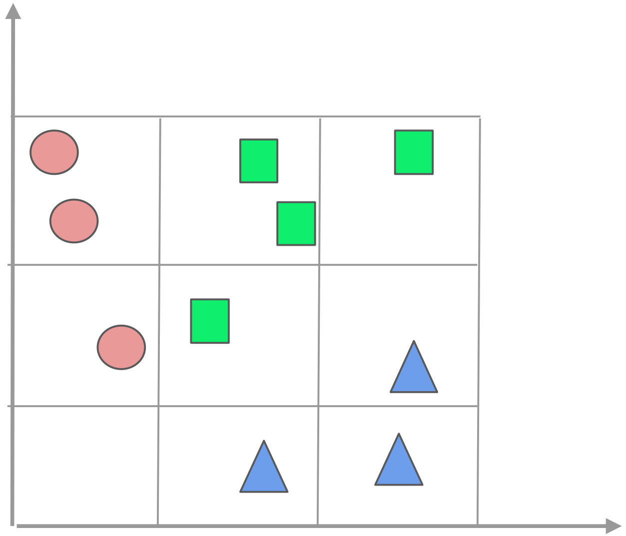
ML models use Statistics



10 Samples, 1 Dimension

How many regions?

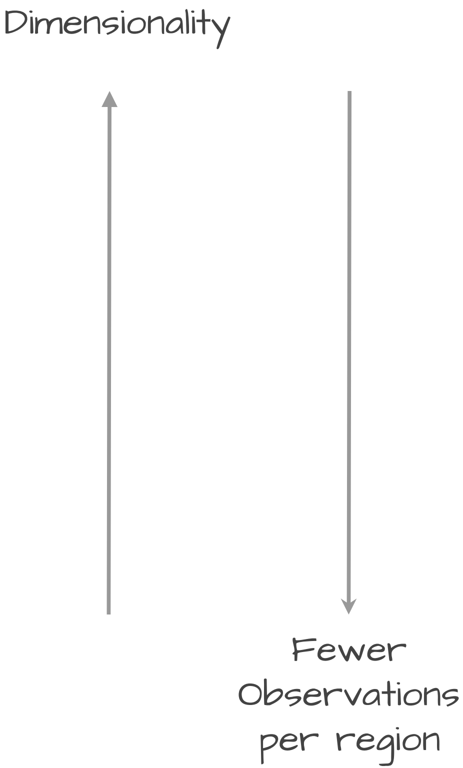
ML models use Statistics



10 Samples, 2 Dimensions

ML models use Statistics

10 Samples, 3 Dimensions, 27 Regions


Dimensionality
Fewer
Observations
per region

How to deal with High dimensional Data?

- Use Domain Expertise
    - Feature Engineering
- Reduce Dimensionality of Data
    - Keep fewer features or Create new dimensions

Dimensionality Reduction

Represent each Sample with fewer Variables

Preserve most of the information in the Data

- Feature Selection
    - Pick a Subset of Original features based on some criteria

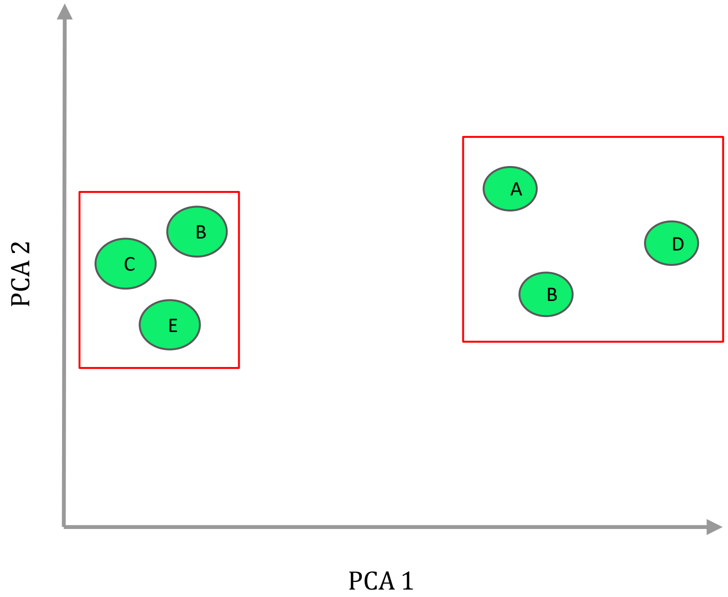
        - e.g Information gain
- Feature Extraction
    - Create completely new set of Features e.g PCA

Principal Component Analysis (PCA)

	F1	F2	F3	F4
A	11	6	12	5
B	10	4	9	7
C	8	5	10	6
D	3	3	2	2
E	2	3	1	4
F	1	2	2	7

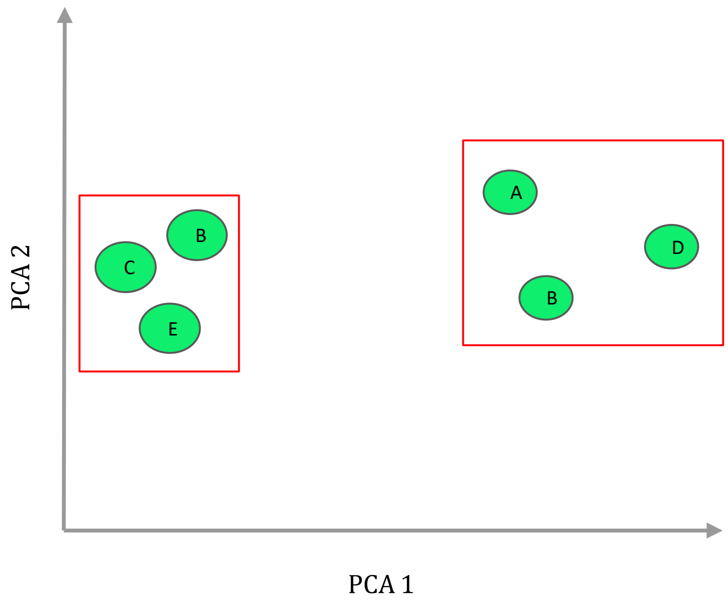
How can PCA help?

	F1	F2	F3	F4
A	11	6	12	5
B	10	4	9	7
C	8	5	10	6
D	3	3	2	2
E	2	3	1	4
F	1	2	2	7



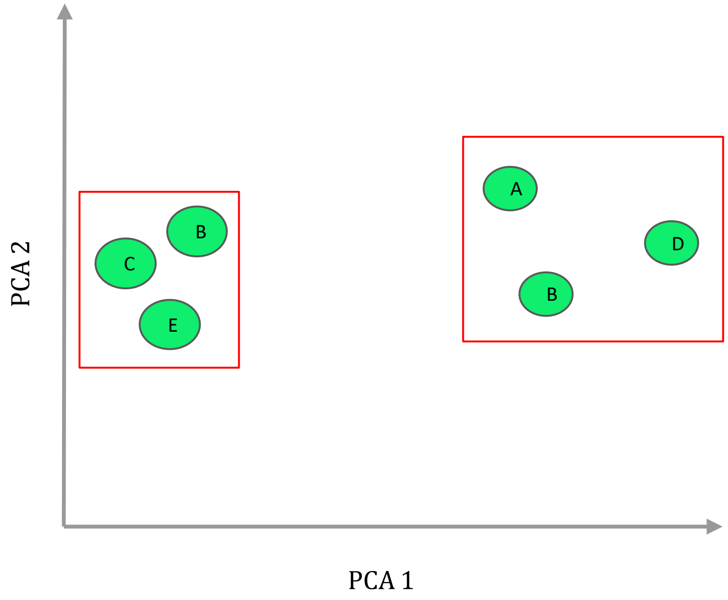
Can make 2D Graph

	F1	F2	F3	F4
A	11	6	12	5
B	10	4	9	7
C	8	5	10	6
D	3	3	2	2
E	2	3	1	4
F	1	2	2	7



Which Feature is most important separating Examples

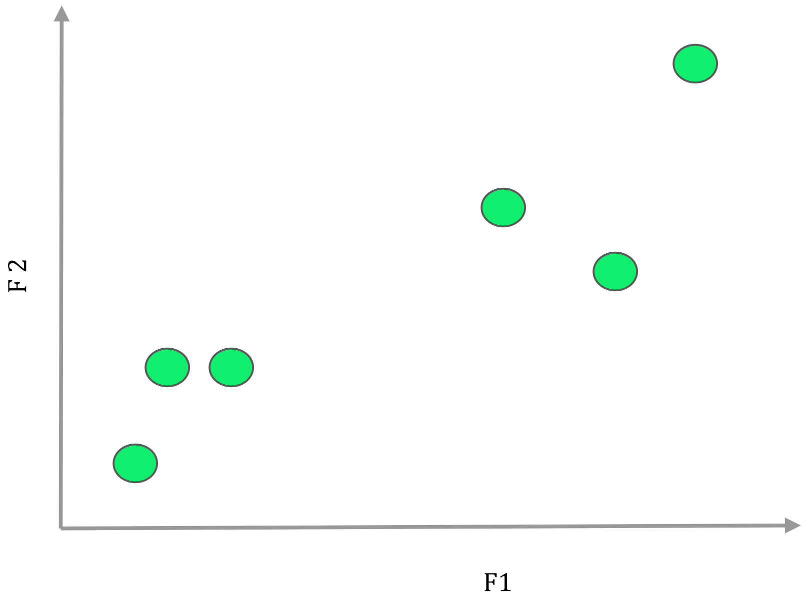
	F1	F2	F3	F4
A	11	6	12	5
B	10	4	9	7
C	8	5	10	6
D	3	3	2	2
E	2	3	1	4
F	1	2	2	7



How accurate the Graph is?

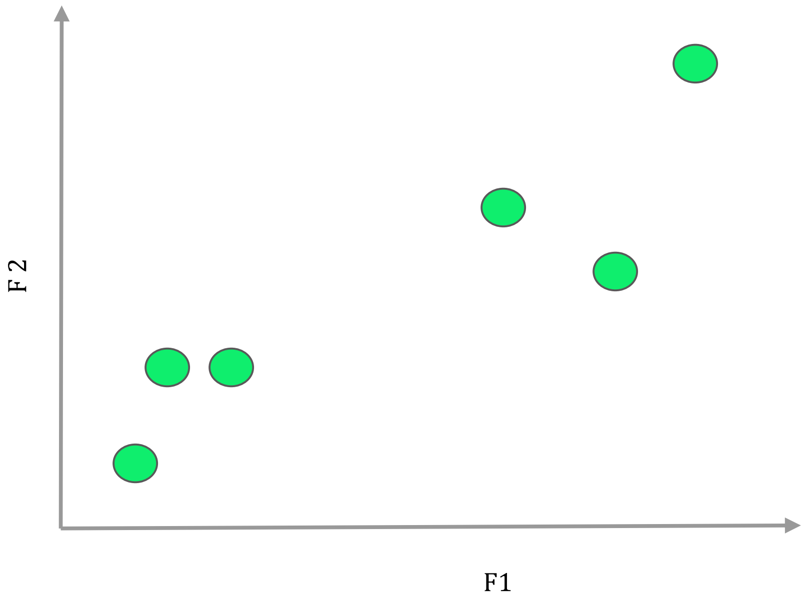
How PCA works?

	F1	F2
A	11	6
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2



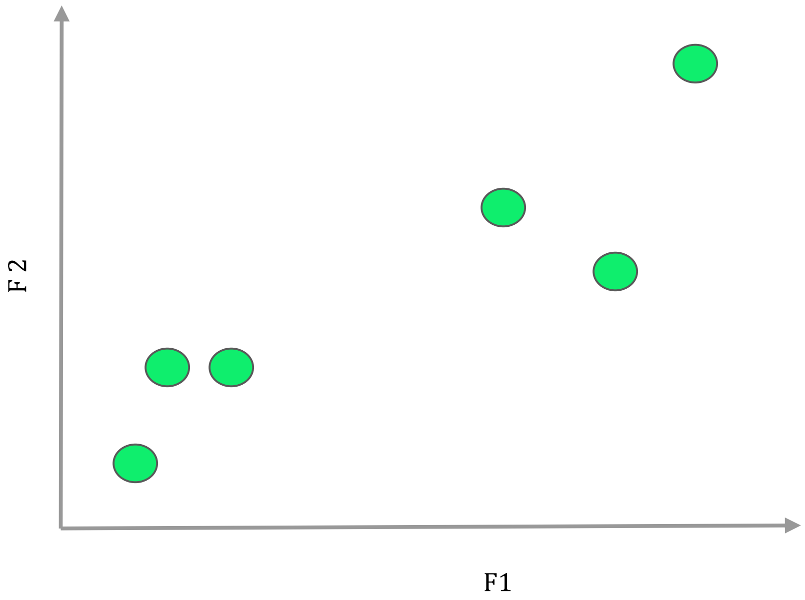
Plot the Graph

	F1	F2
A	11	6
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2



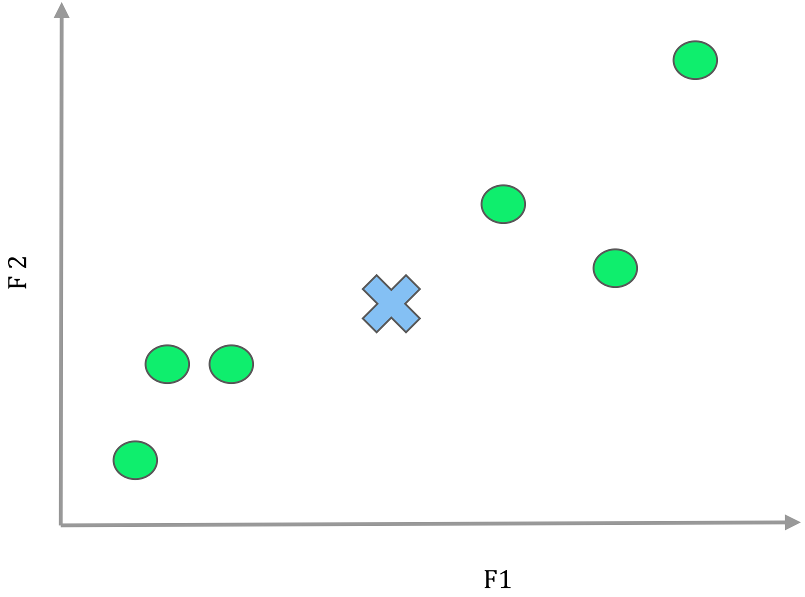
Calculate feature(s) mean

	F1	F2
A	11	6
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2
Mean	5.83	3.83



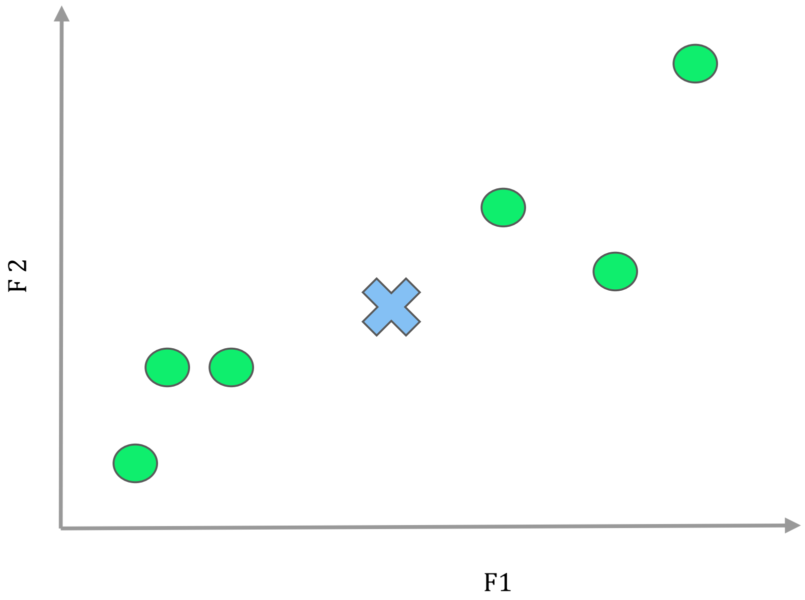
Calculate feature(s) mean

	F1	**F2**
A	11	6
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2
Mean	5.83	3.83

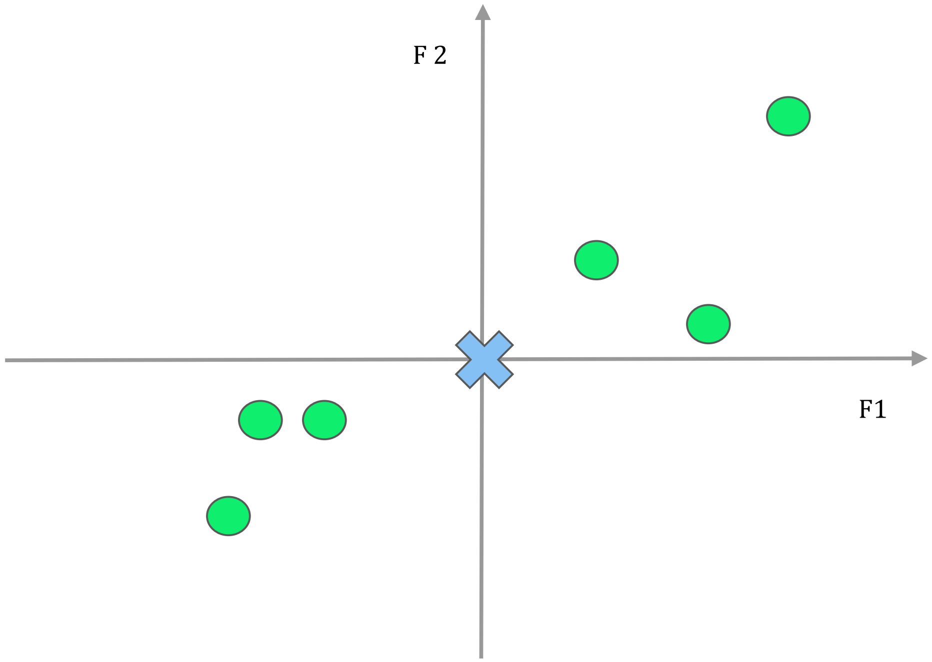


Plot the mean

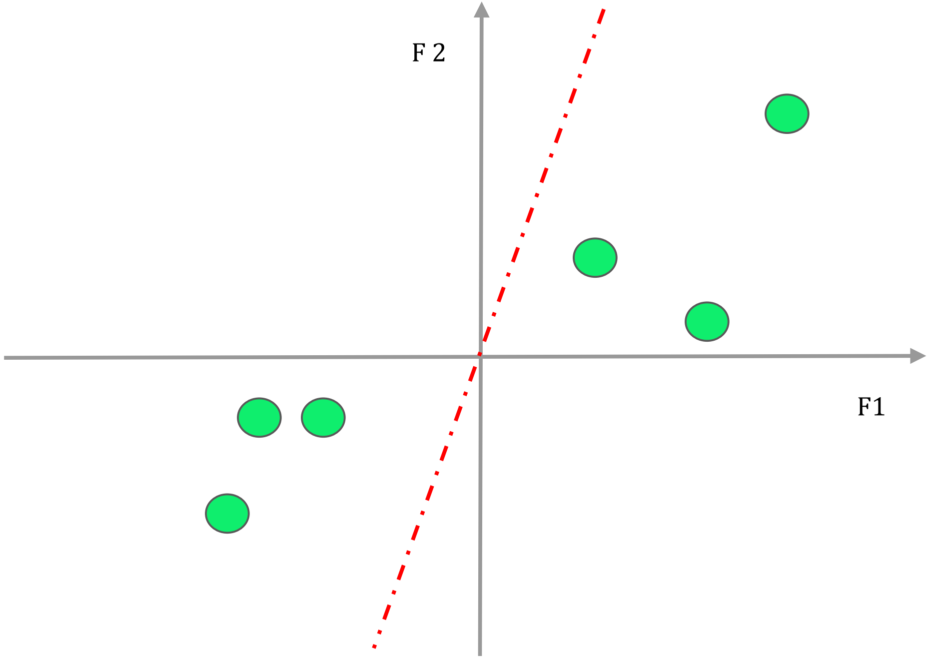
	F1	F2
A	11	6
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2
Mean	5.83	3.83



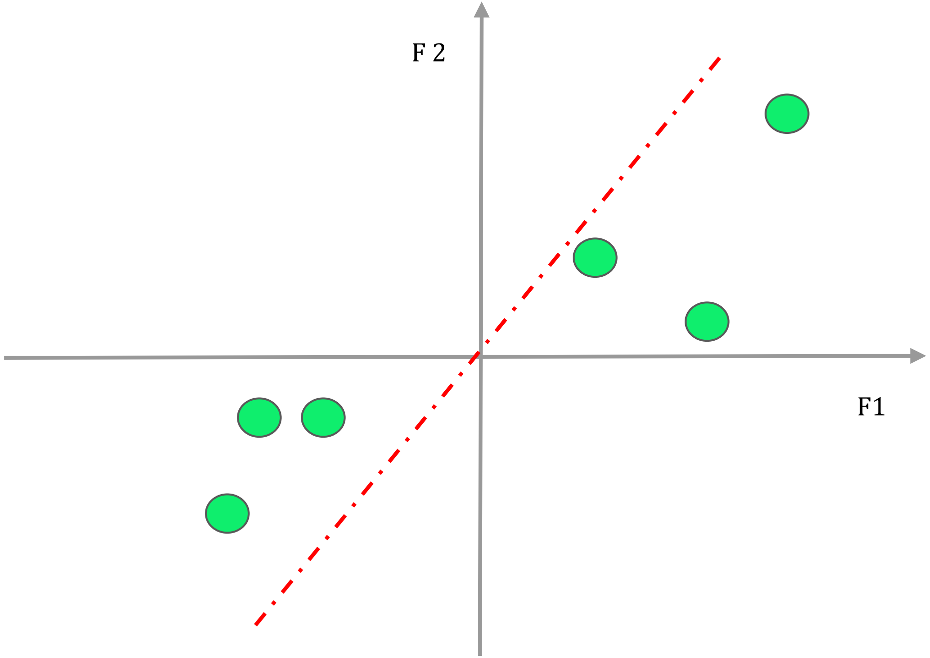
Shift the Origin to Mean



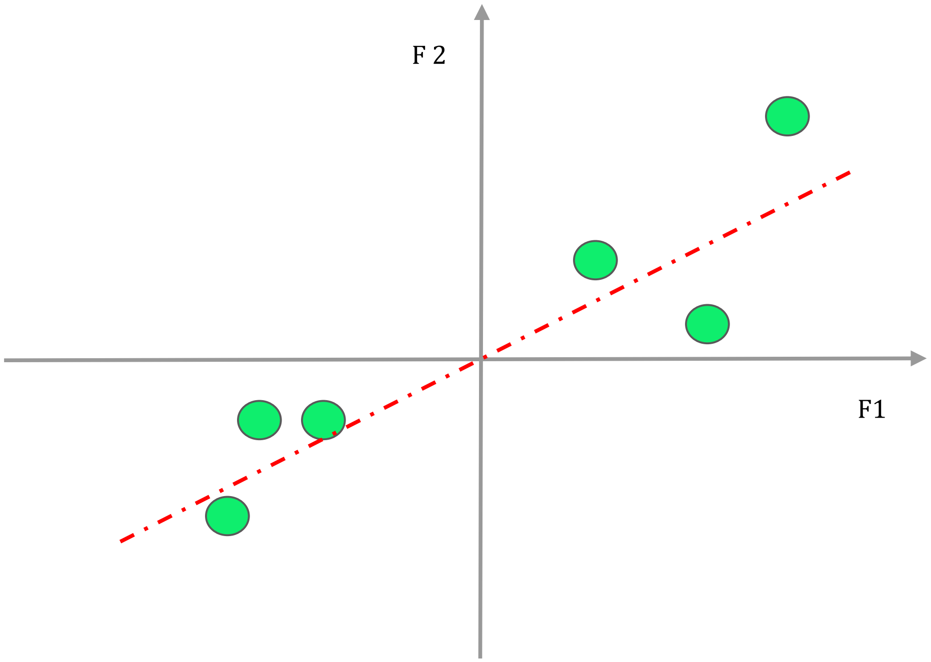
Shift the Origin to Mean



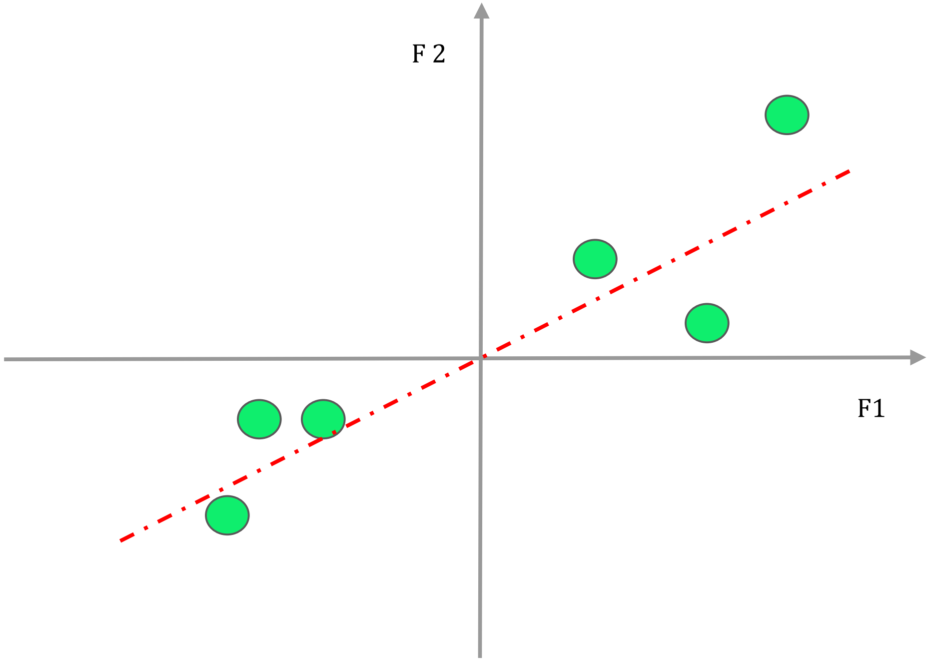
Draw a line through Origin



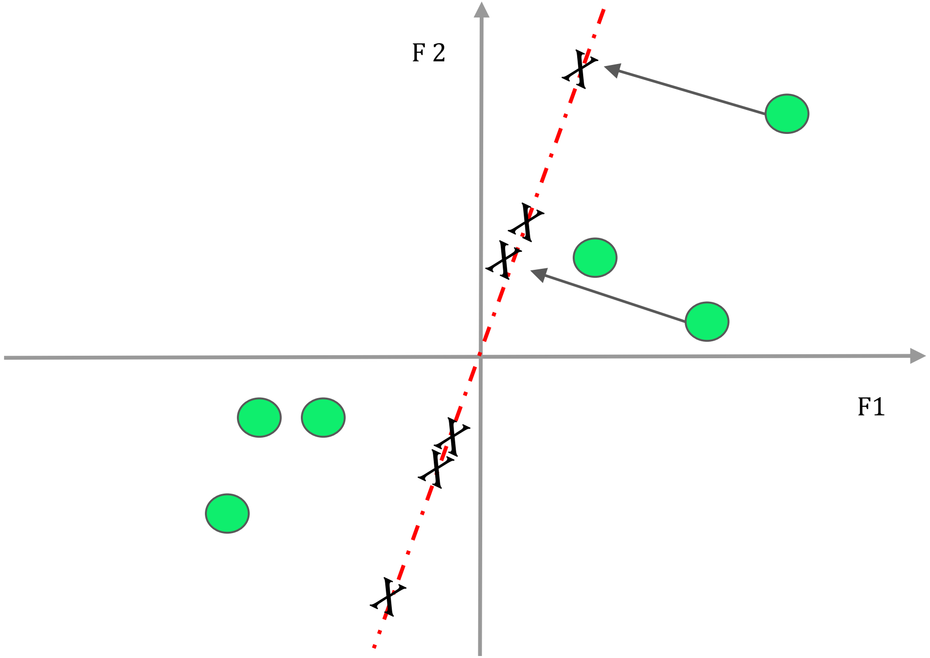
Rotate the line



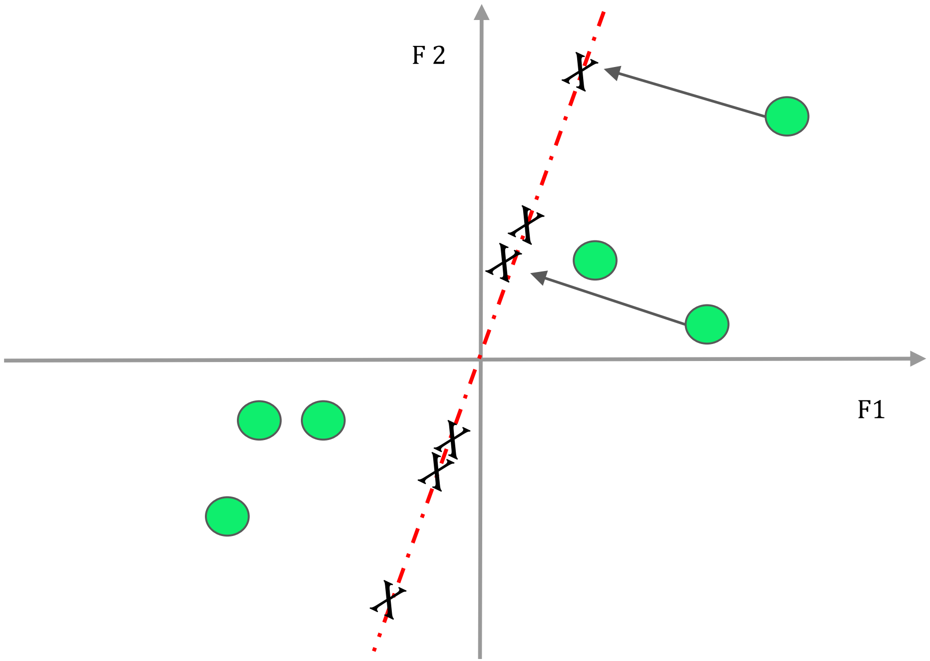
Keep rotating till best fit is found



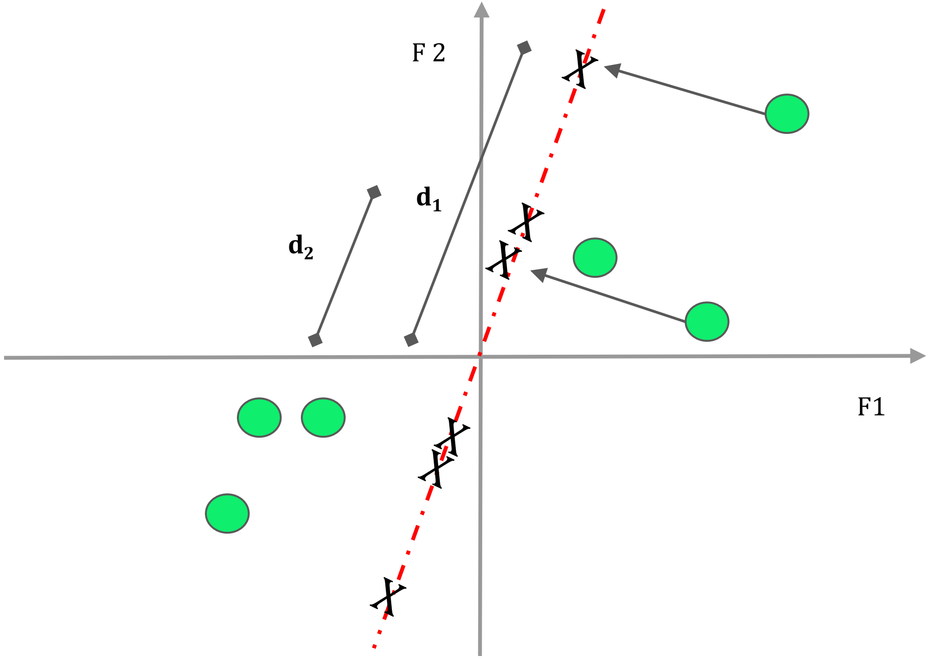
What does best fit here?



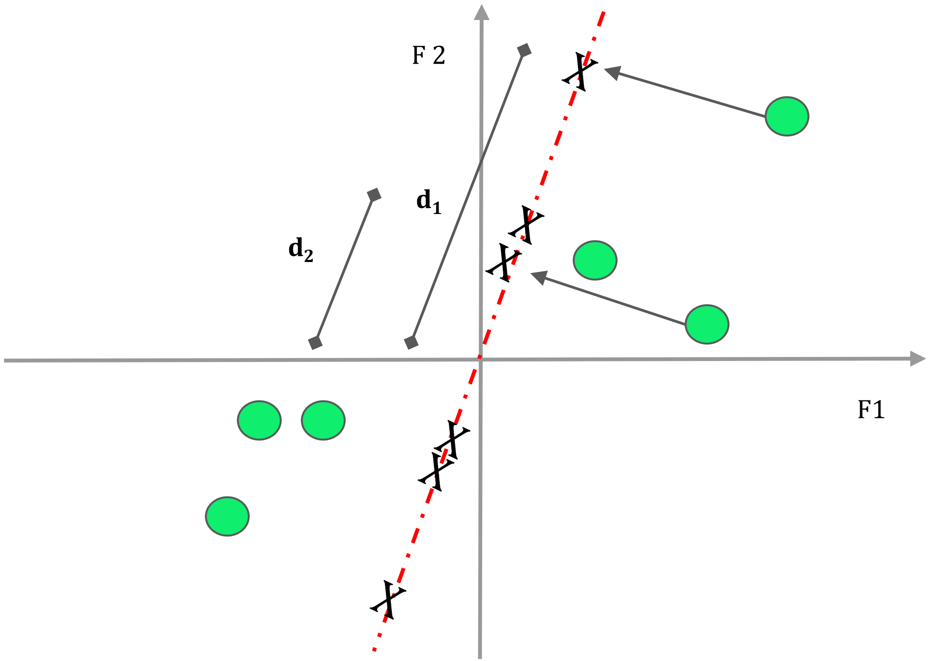
Project Data points on the line



Rotate line until Projections are minimum

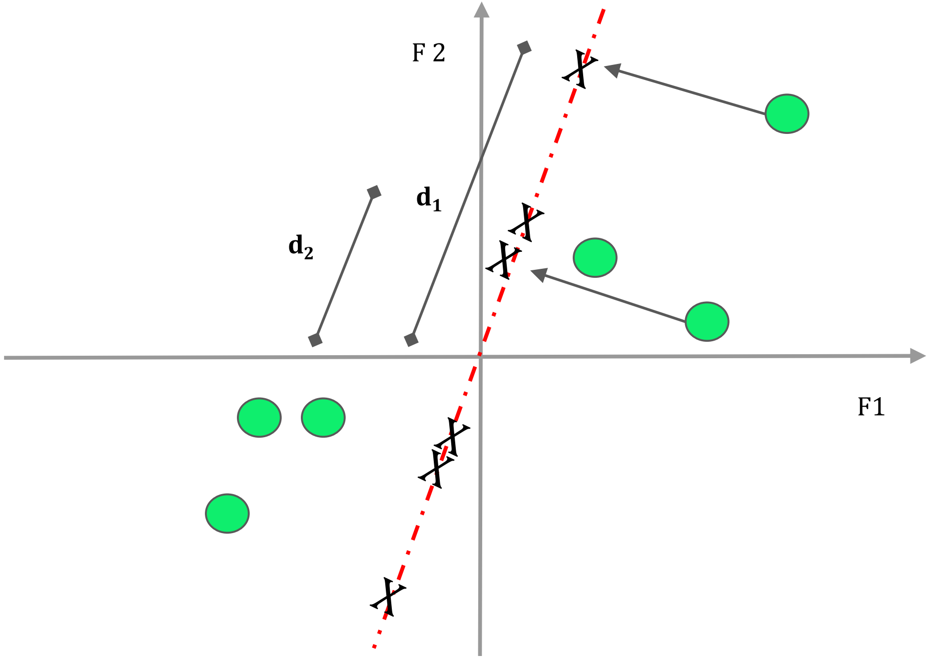


OR distance from Origin is Maximum

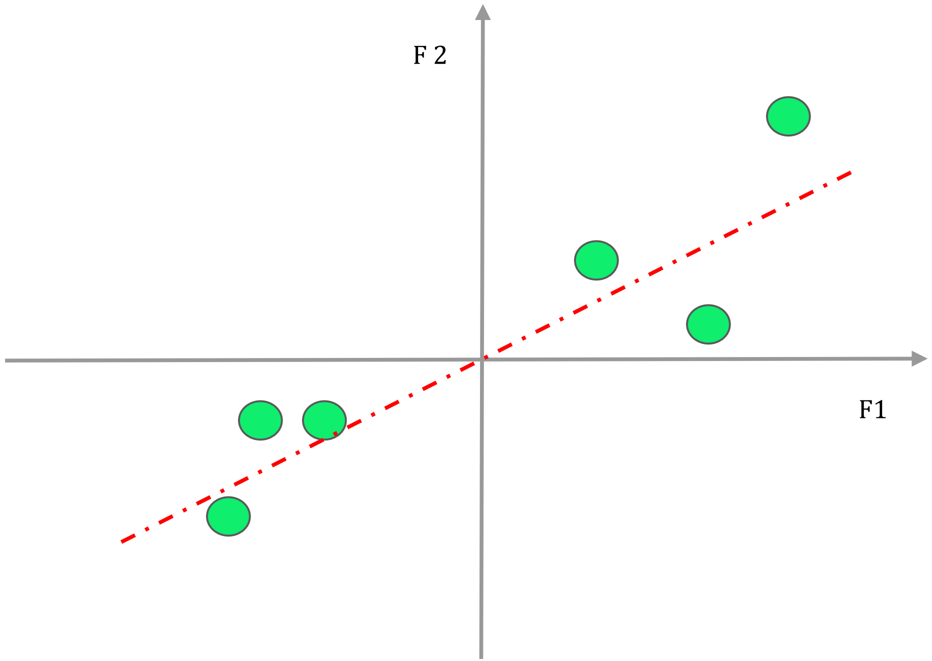


How do we compare distances?

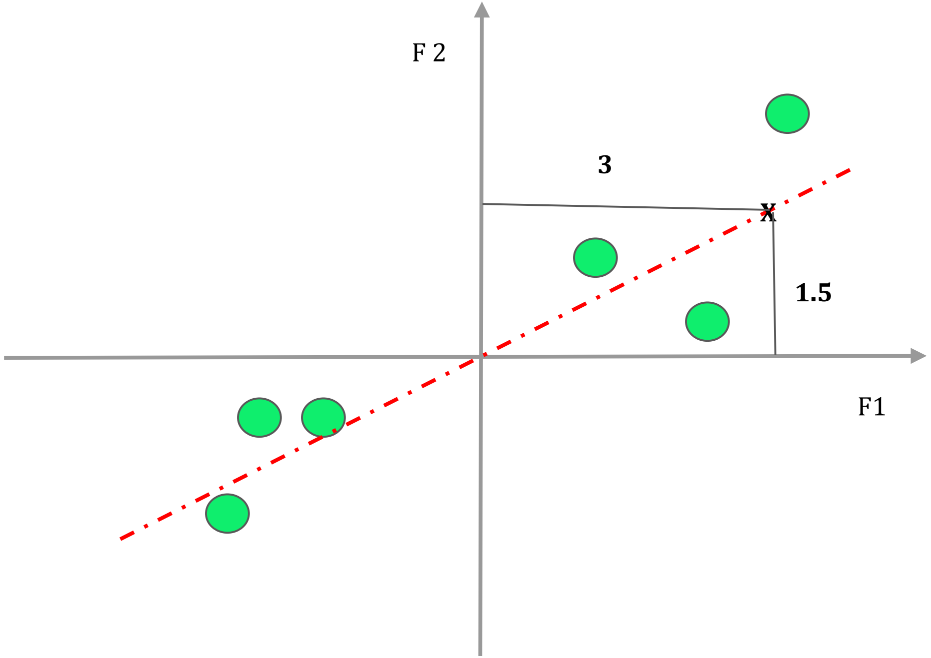
Sum of Square distances = $d_1{}^2 + d_2{}^2 + d_3{}^2 + d_4{}^2 + d_5{}^2 + d_6{}^2$



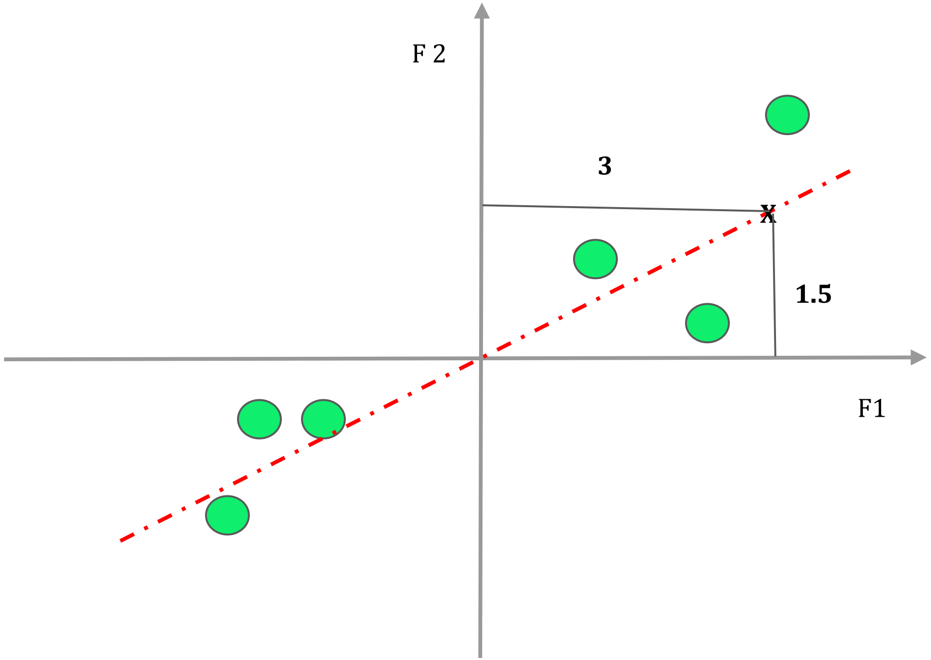
How do we compare distances?



Best fit line is called 'PC 1'



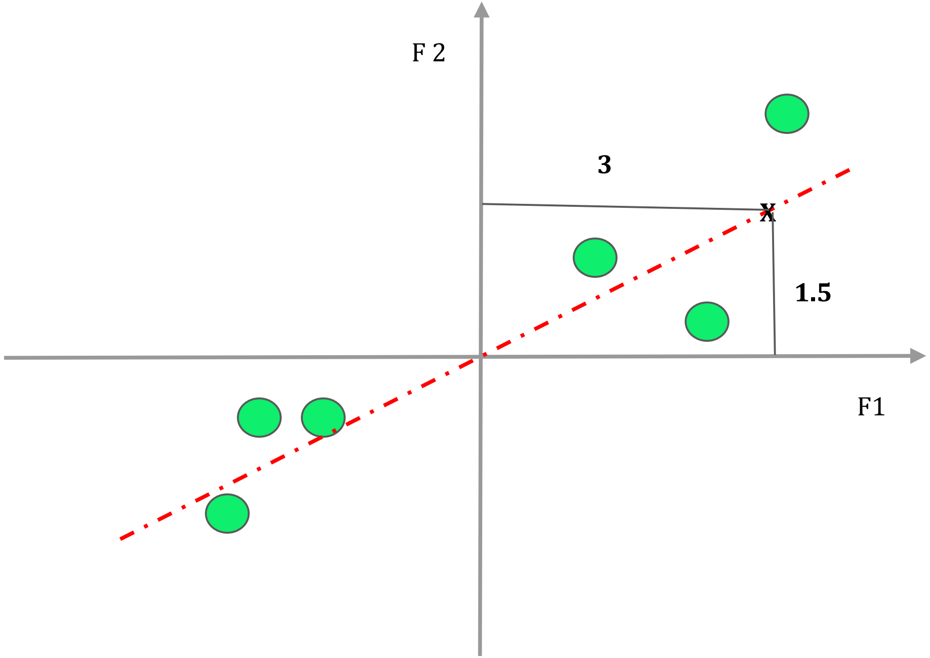
Calculate Ratio of F1 and F2 for Best Fit line



F1 is 2 times more important than F2

3

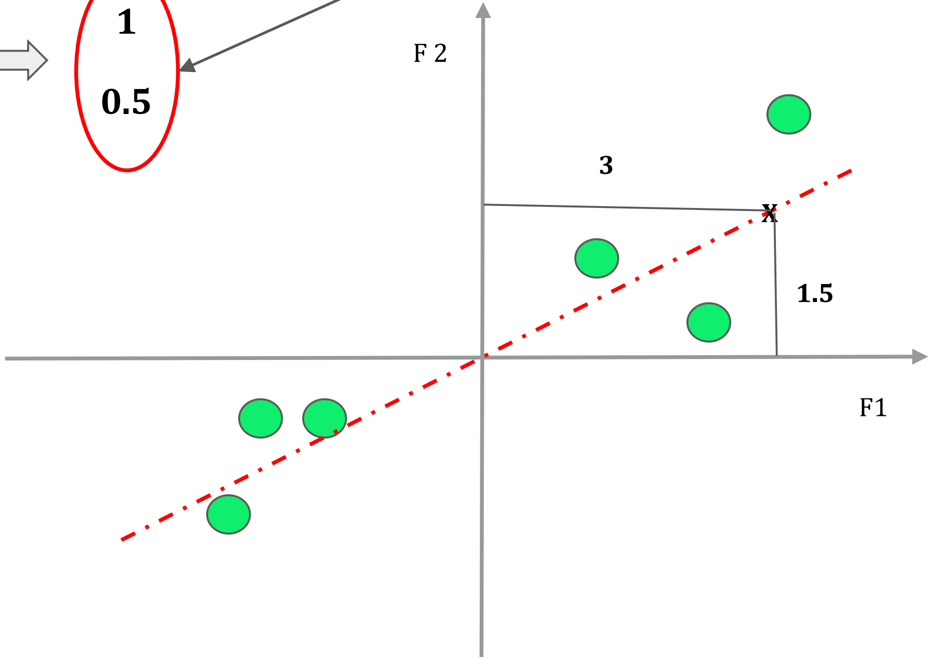
1.5



F1 is 2 times more important than F2

Eigenvector for PC 1

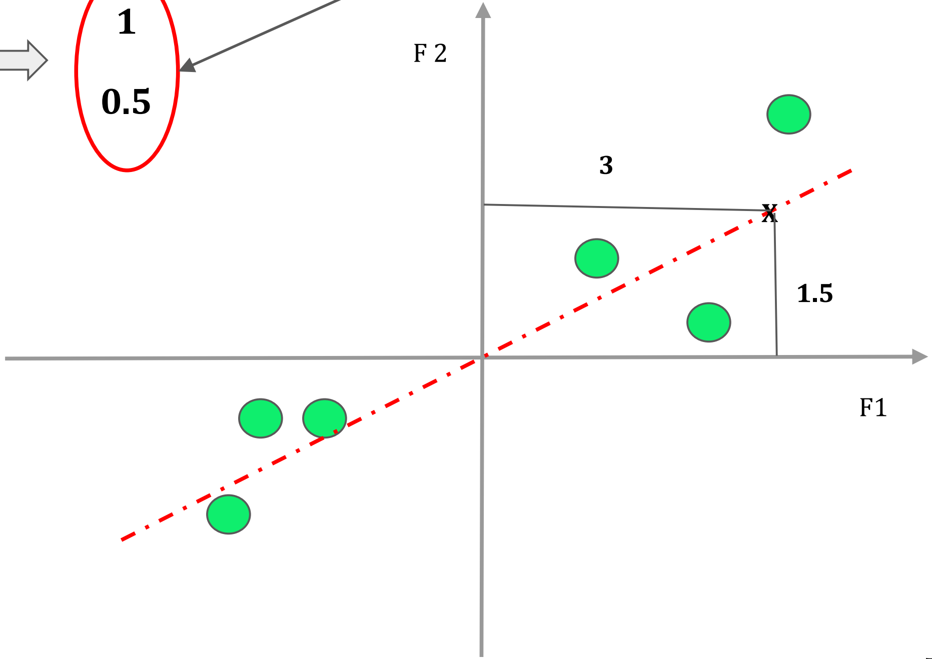
$$\begin{bmatrix} 3 \\ 1.5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$



Make it a Unit Vector

EigenVector for PC 1

$$\begin{matrix}3\\1.5\end{matrix} \Rightarrow \begin{matrix}1\\0.5\end{matrix}$$



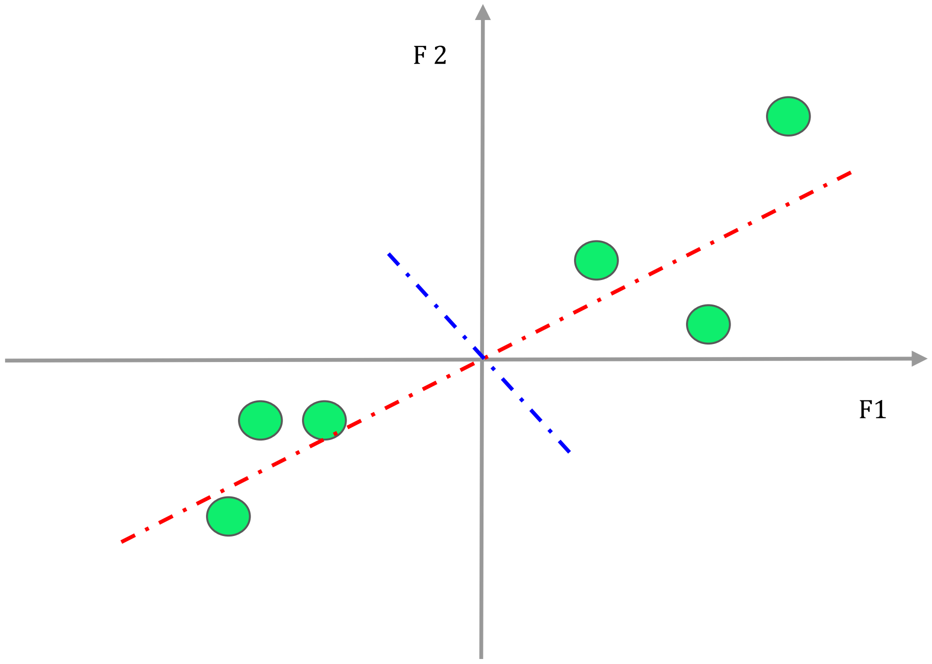
EigenValue for PC 1

Sum of Square distances = $\mathbf{d_1^{\,2} + d_2^{\,2} + d_3^{\,2} + d_4^{\,2} + d_5^{\,2} + d_6^{\,2}}$

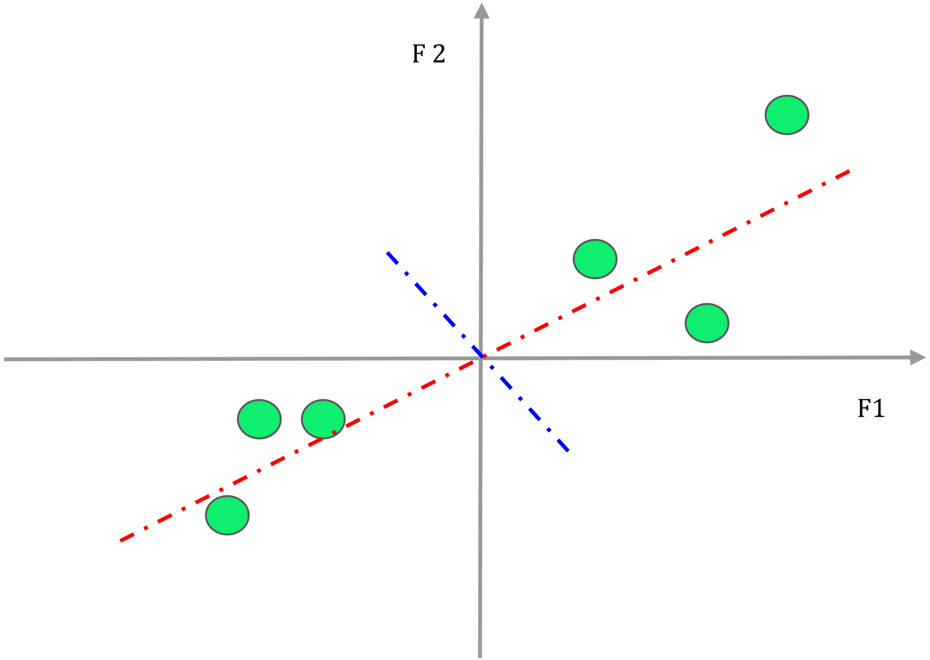
So far ...

1. Calculate mean of each feature
2. Subtract mean (Move origin)
3. Draw random line and Project points
4. Calculate Sum of Square distances from Origin to Projection
5. Rotate line for Best fit (maximum SSD) - **PC 1**
6. Find Feature ratio to get Feature importance
7. Calculate Unit Vector - **EigenVector for PC 1**
8. Find SSD for PC1 - **EigenValue of PC1**

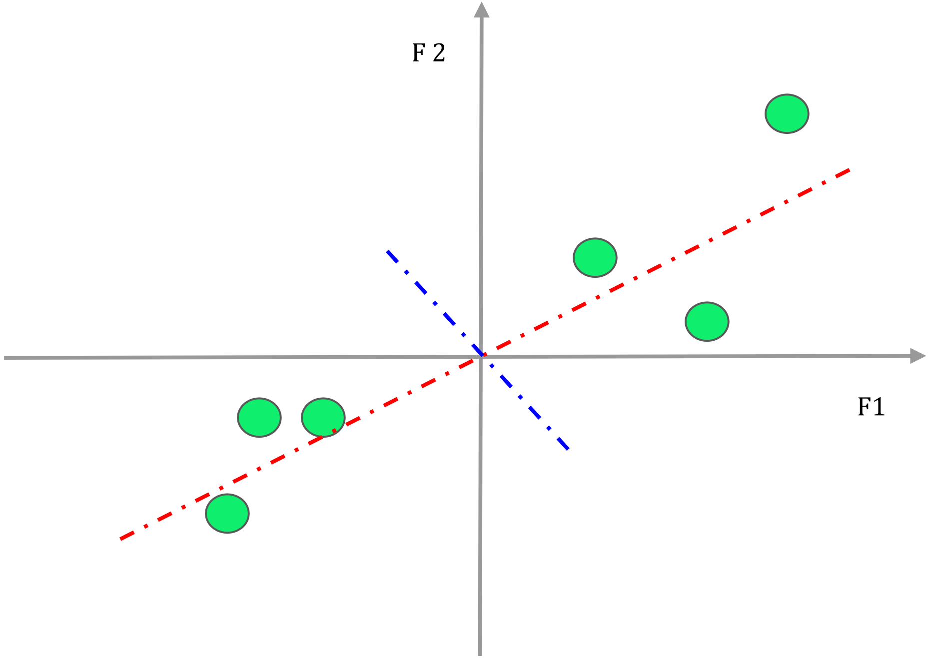
What’s Next?



Draw a Perpendicular line to PC 1



This is PC 2



Calculate EigenValue and EigenVector for PC2

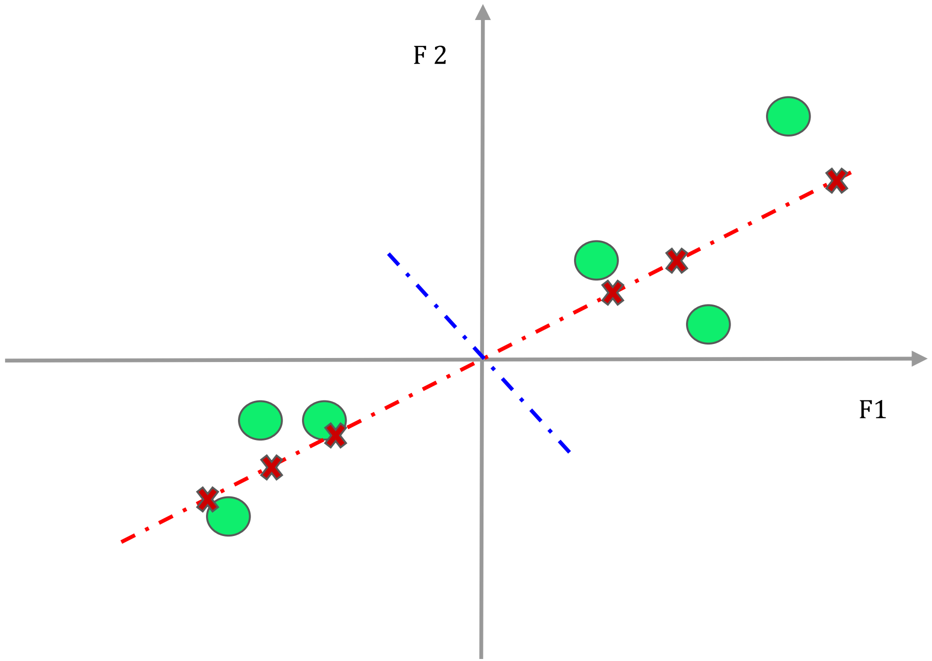
What’s the role of EigenValue?

Variance in the original data captured by PC 1

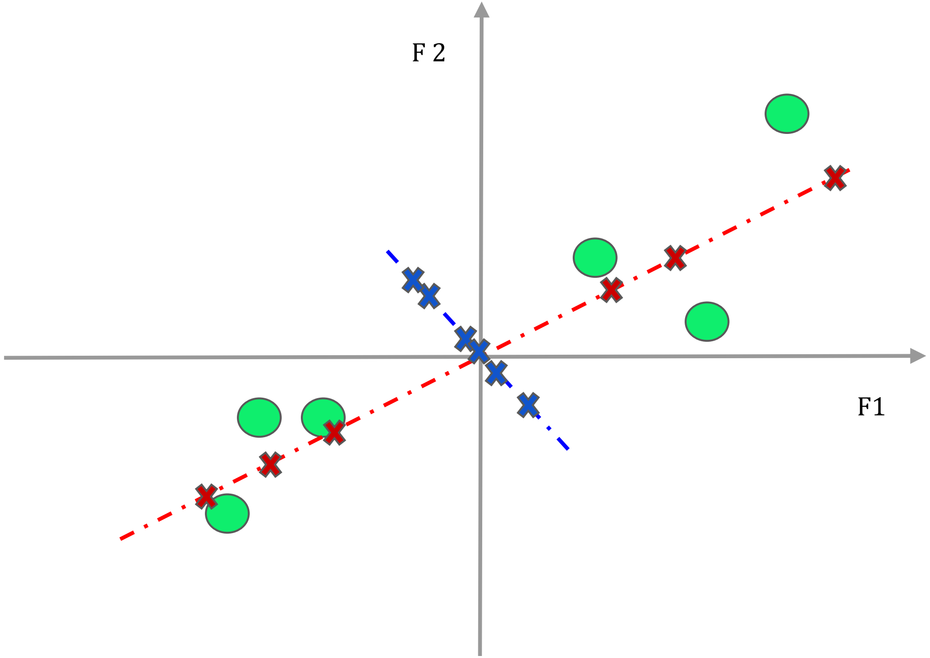
PC 1	**18**	**85.5%**
PC 2	**3**	**14.5%**

Variance in the original data captured by PC 2

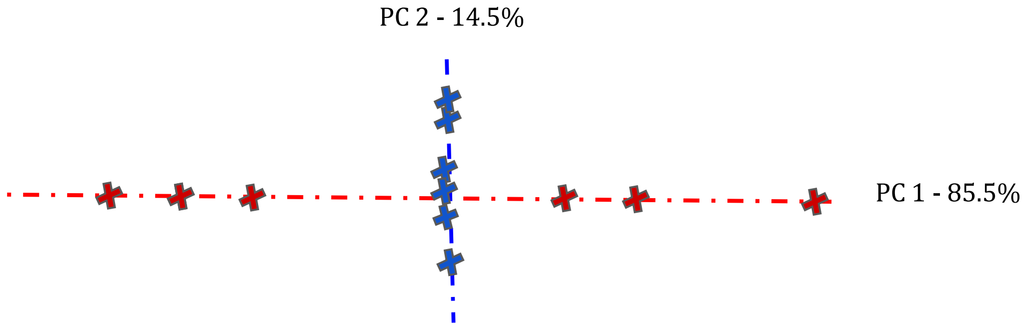
What are the new Dimensions or Features?



PC 1 Dimension



PC 2 Dimension



PC1, PC2 are new Dimensions

Dimensionality Reduction

	F1	F2	F3
A	11	6	12
B	10	4	9
C	8	5	10
D	3	3	2
E	2	3	1
F	1	2	2

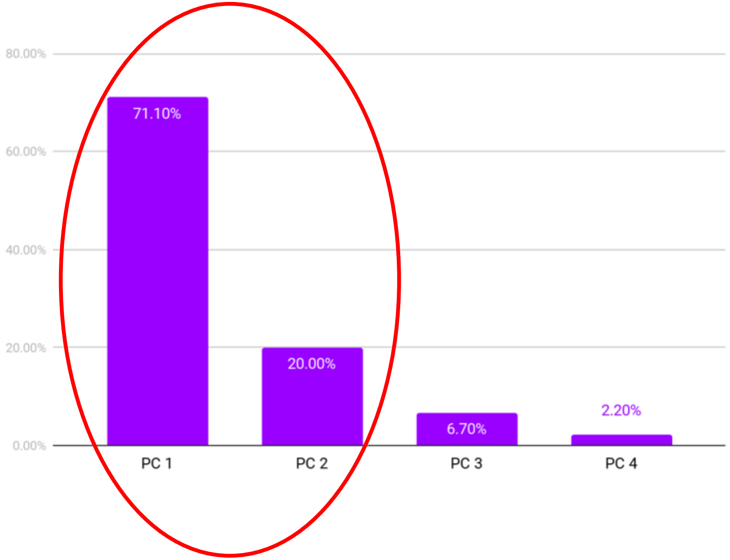
	F1	F2	F3	F4
A	11	6	12	5
B	10	4	9	7
C	8	5	10	6
D	3	3	2	2
E	2	3	1	4
F	1	2	2	7

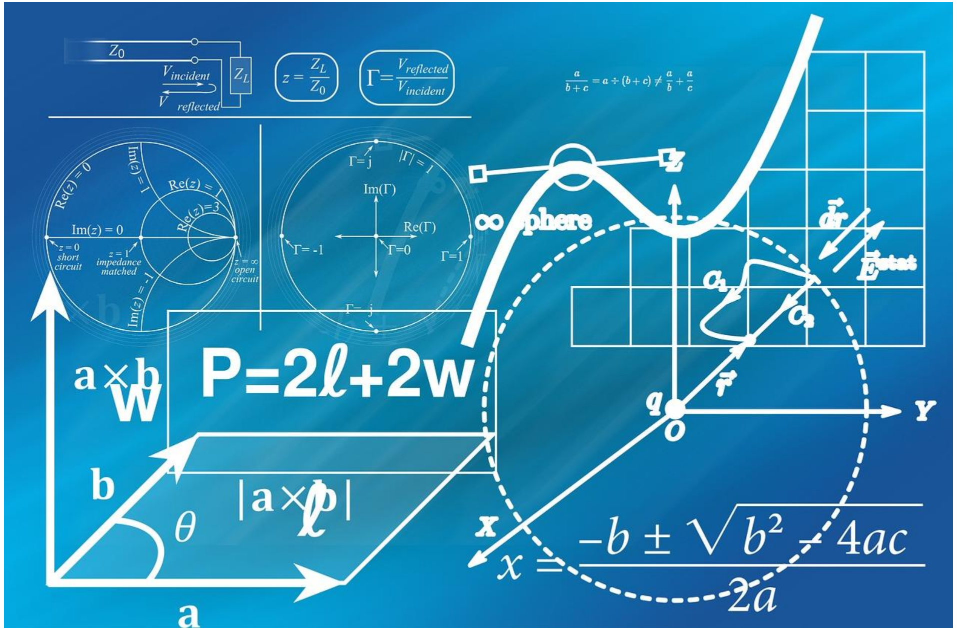
Calculate EigenVectors

Calculate EigenValues

How many EigenVectors and EigenValues?

	EigenValue	%Variance
PC1	32	71.1%
PC2	9	20.0%
PC3	3	6.7%
PC4	1	2.2%





P=2ℓ+2w
a×b
|a×b|
b
a
θ
Im(z) = 0
Im(Γ)
Re(Γ)
Γ=0
Γ= -1
Γ=1
short circuit
impedance matched
open circuit
x
Y
O
q
$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Math of PCA

	F1	F2
A	12	7
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2

$$\begin{bmatrix} 12 & 7 \\ 10 & 4 \\ 2 & 3 \\ 8 & 5 \\ 3 & 3 \\ 1 & 2 \end{bmatrix}$$

The Matrix

	F1	F2
A	12	7
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2
Mean	**6**	**4**

$$\begin{bmatrix} 12-6 & 7-4 \\ 10-4 & 4-4 \\ 2-6 & 3-4 \\ 8-6 & 5-4 \\ 3-6 & 3-4 \\ 1-6 & 2-4 \end{bmatrix}$$

Calculate Mean

	F1	F2
A	12	7
B	10	4
C	2	3
D	8	5
E	3	3
F	1	2
Mean	**6**	**4**

$$\begin{bmatrix} 6 & 3 \\ 4 & 0 \\ -4 & -1 \\ 2 & 1 \\ -3 & -1 \\ -5 & -2 \end{bmatrix}$$

Mean after shift?

Variance & Covariance

Variance

Measure of spread of data in feature...

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

Covariance

How two features move together in a dataset...

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

$$A = \begin{bmatrix} 6 & 3 \\ 4 & 0 \\ -4 & -1 \\ 2 & 1 \\ -3 & -1 \\ -5 & -2 \end{bmatrix}$$

Using Matrix Multiplication to get Covariance Matrix

$$Cov(x,y,z) = \begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

$$Cov(x,y) = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix}$$

$$A = \begin{bmatrix} 6 & 3 \\ 4 & 0 \\ -4 & -1 \\ 2 & 1 \\ -3 & -1 \\ -5 & -2 \end{bmatrix}$$

$$cov(x, y) = A^T . A$$

$$Cov(x, y) = \begin{bmatrix} 6 & 4 & -4 & 2 & -3 & -5 \\ 3 & 0 & -1 & 1 & -1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 6 & 3 \\ 4 & 0 \\ -4 & -1 \\ 2 & 1 \\ -3 & -1 \\ -5 & -2 \end{bmatrix}$$

$$cov = \begin{bmatrix} 106 & 37 \\ 37 & 106 \end{bmatrix}$$

$$cov = \begin{bmatrix} 21.2 & 7.4 \\ 7.4 & 21.2 \end{bmatrix}$$

EigenVector and EigenValues of Covariance Matrix

$$Ae = \lambda e$$

Covariance Matrix

EigenValue

EigenVector

$$Ae = \lambda I e$$

$$Ae - \lambda I e = 0$$

$$(A - \lambda I)e = 0$$

$$det(A - \lambda I) = 0$$

$$\begin{bmatrix} 21.2 & 7.4 \\ 7.4 & 21.2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 21.2 & 7.4 \\ 7.4 & 21.2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$\begin{bmatrix} 21.2 - \lambda & 7.4 \\ 7.4 & 21.2 - \lambda \end{bmatrix}$$

$$det(A - \lambda I) = 0$$

$$(21.2 - \lambda)(21.2 - \lambda) - (7.4)(7.4) = 0$$

$$\lambda^2 - 42.4\lambda + 449.44 - 54.76 = 0$$

$$\lambda^2 - 42.4\lambda + 394.68 = 0$$

$$\lambda^2 - 42.4\lambda + 394.68 = 0$$

How to Solve this :)

$$(\lambda - 28.6)(\lambda - 13.8) = 0$$

$\lambda = 28.6$ and $\lambda = 13.8$

What does λ represent?

EigenVector or Principal Components

$$(A - \lambda I)e = 0$$

$$\begin{bmatrix} 21.2 - \lambda & 7.4 \\ 7.4 & 21.2 - \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 21.2 - 28.6 & 7.4 \\ 7.4 & 21.2 - 28.6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -7.4 & 7.4 \\ 7.4 & -7.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 21.2 - 13.8 & 7.4 \\ 7.4 & 21.2 - 13.8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 7.4 & 7.4 \\ 7.4 & 7.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Improving Classification with PCA

Dimensionality Reduction with PCA

Noise Reduction with PCA

Dimensionality Reduction

Advantages ...

- Allows working with High Dimensional Data
- Reduced Dataset size
- Faster Processing

Disadvantages...

- Computationally Expensive for Large feature space
- Linear combination will struggle to handle non-linear relationship between features