Regularization

— avoid overfitting

Reason for overfitting

— too many irrelevant features

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \qquad + \beta_n x_n$$

CRIM INDUS CHAS NOX

1 Remove some irrelevent features (L1 Reg)

2 Reduce the importance of " " (L2 Reg)

✓ Lasso → Linear Reg + L1 Reg

— Value of β's are reduced

— sometimes they become zero

✓ Ridge → Linear Reg + L2 Reg

— values of β's are reduced

— But they dont become zero

lasso = Lasso( )

lasso.fit(X, y)

lasso.coef_

— the β's or model parameters

Decision Tree →

→ restrict the depth

✓ ↳ max_depth = 6

✓ leaves → 10

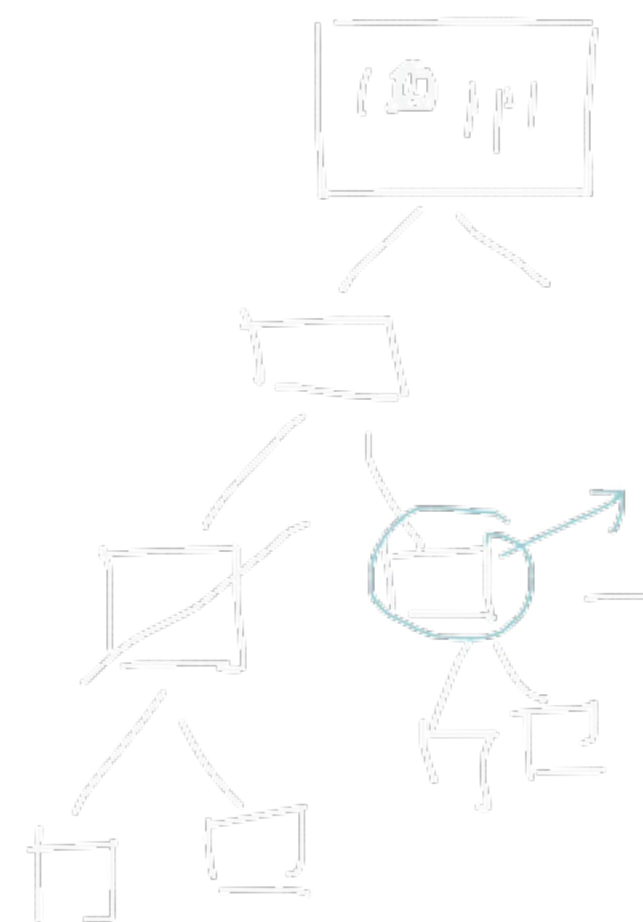
✓ min samples →

depth = 10

— 10 conditions

depth = 64

— 64 conditions



→ Gain < γ

1 pruning → After tree is built

2 . gamma → while the tree ✓

SVM → gamma (controls overfitting)

→ C (tolerance factor)

Dimensionality Reduction

↓

No of Features
in the model

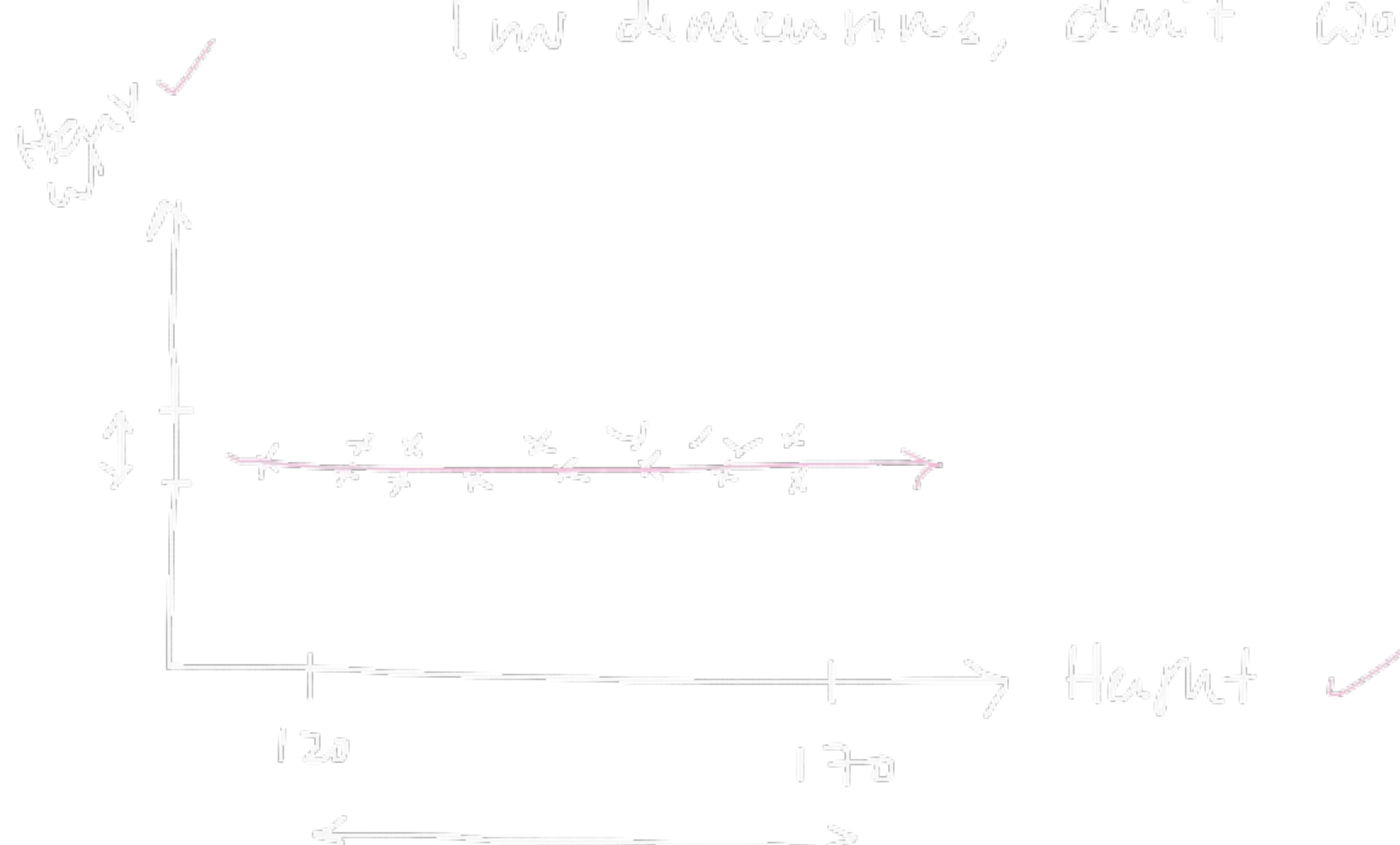
'Curse of dimensionality' → (a) Space & time complexity is high

7000 Feat

(b) Algorithms that work well in low dimensions, don't work well

→ Noise

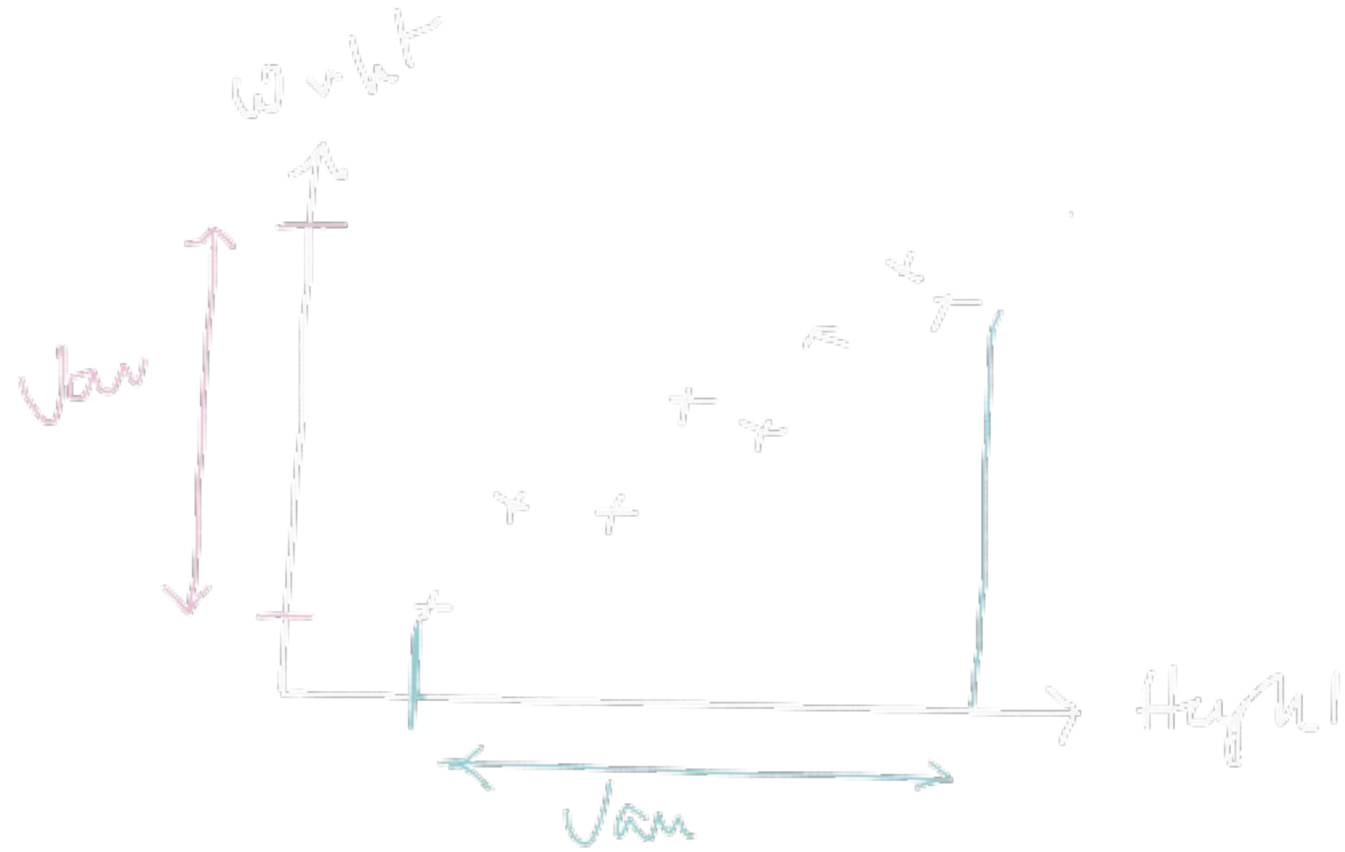
Height	~~Hair colour~~
120	1/0
170	0
99%	1%



50 %	50 %
Height	Wt
120	40
170	100

2d → 1d

15% info loss



Challenge : Reduce no of features without significant loss of information

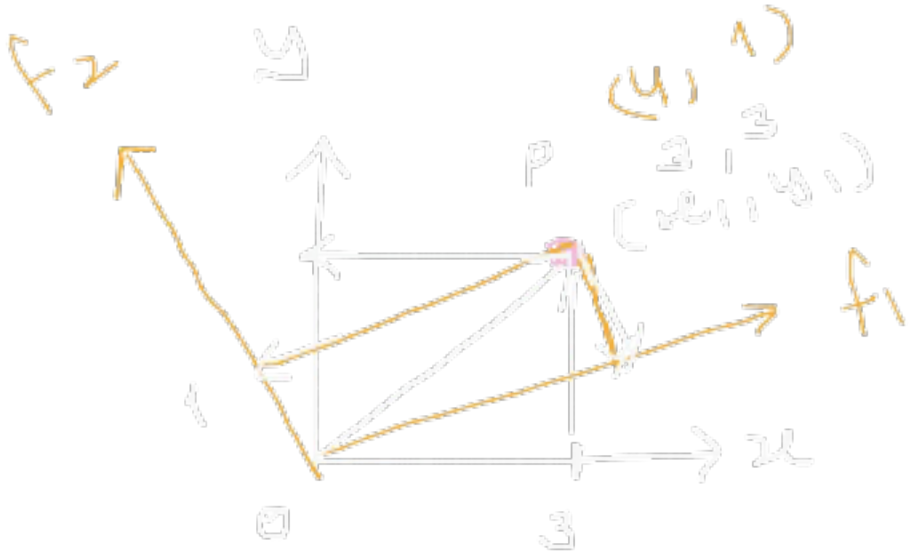
- Principal Component Analysis (PCA)
- t-distributed Stochastic Neighbourhood embedding (t-SNE)

Objective — 1. Make it easy to build models (PCA)
Get rid of the curse of dimensionality

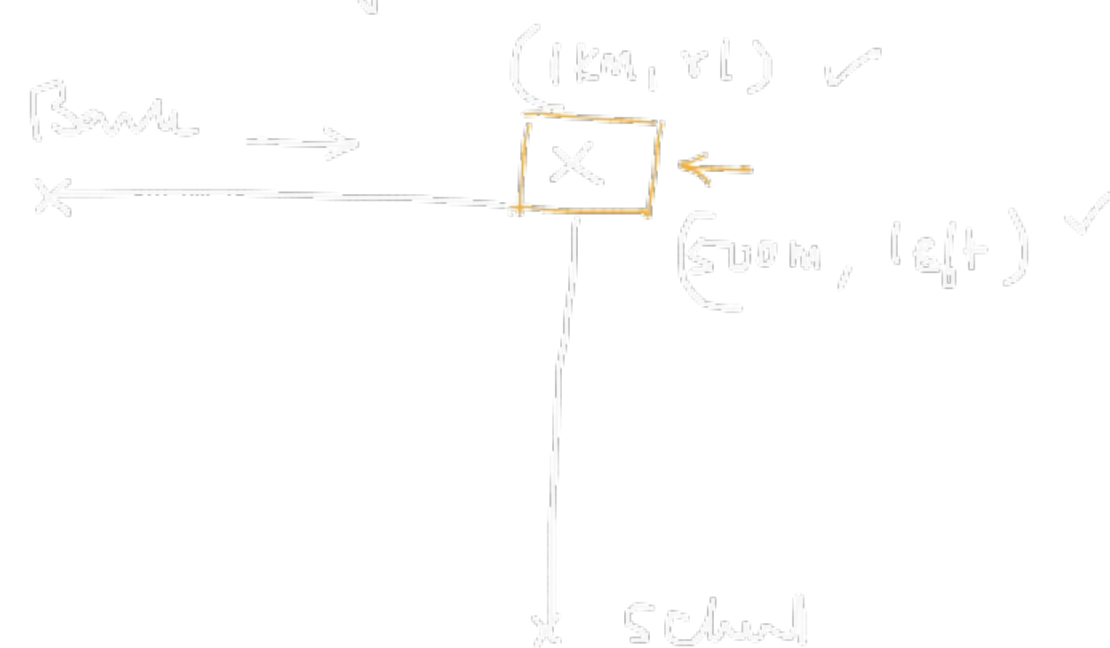
How many columns will we keep?

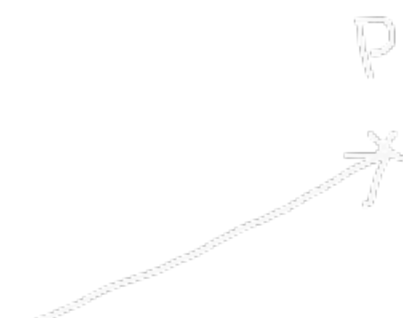
2 Visualize data (t-SNE)
— always 2 columns

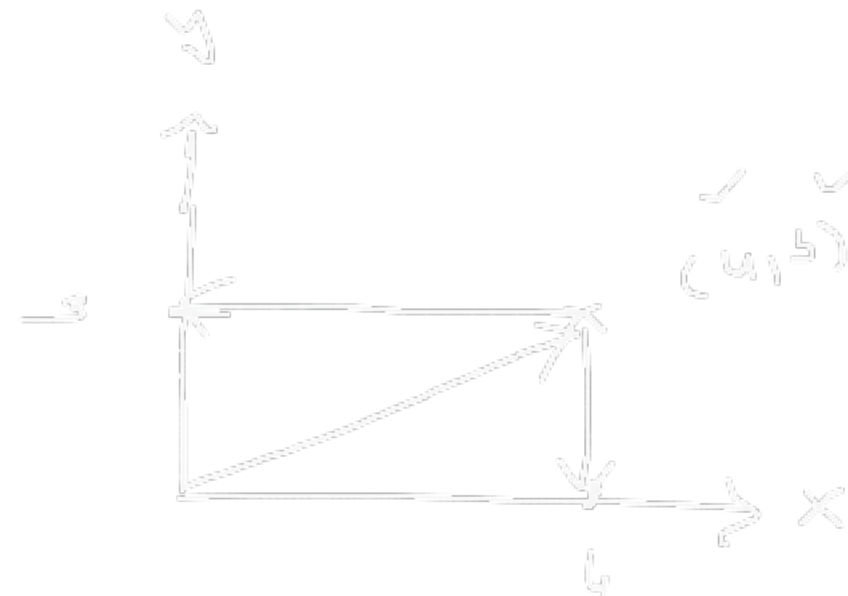
Principal Component Analysis (PCA)

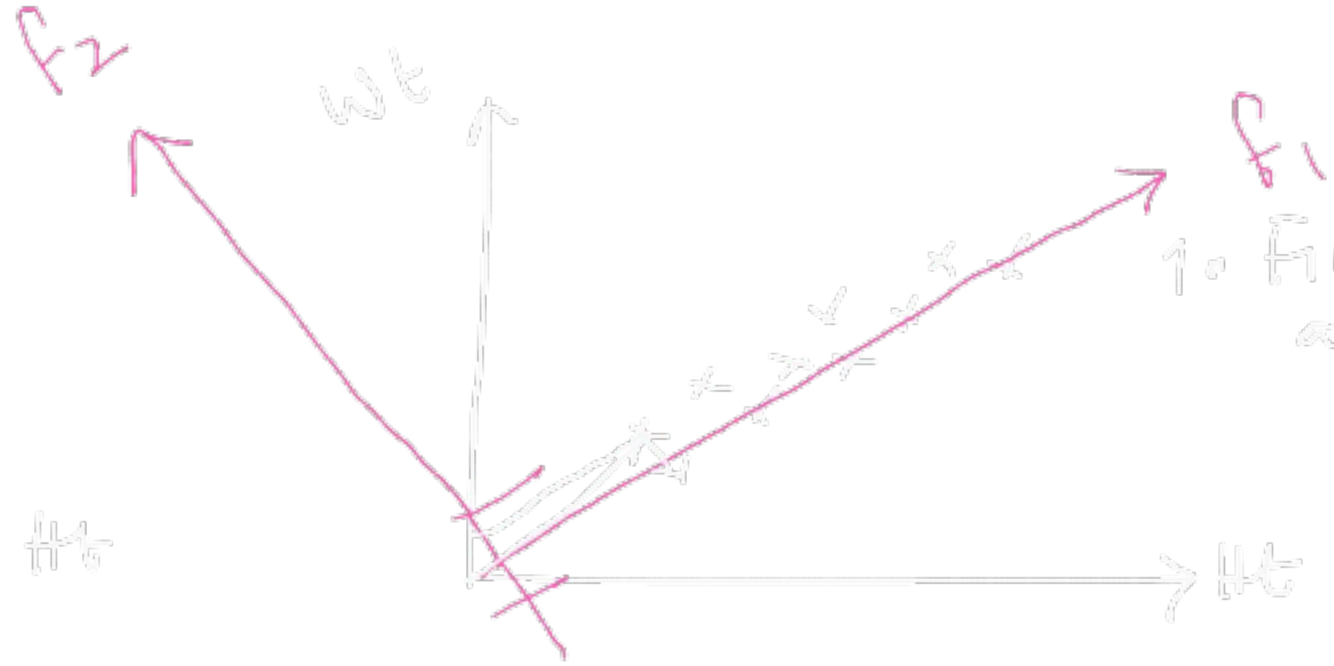


	x	y		f1	f2
P_1	3	3	→	4	1
P_2	4	2	→	3	2
P_3	5,	1	→	4	2









1. Find this direction along which you have max Variance

2 - Translate all points from Ht & Wt → f_1 f_2 coord

2d			1d
Ht	Hc	→	Ht

Ht Wt	Speed

98%

Size	Speed

← 2d →

Ht	Wt	→	f_1 (98%)	F_2 (X 2Y)

↓ Size

$\longleftarrow 3d \longrightarrow$

$x_1 \quad x_2 \quad x_3$

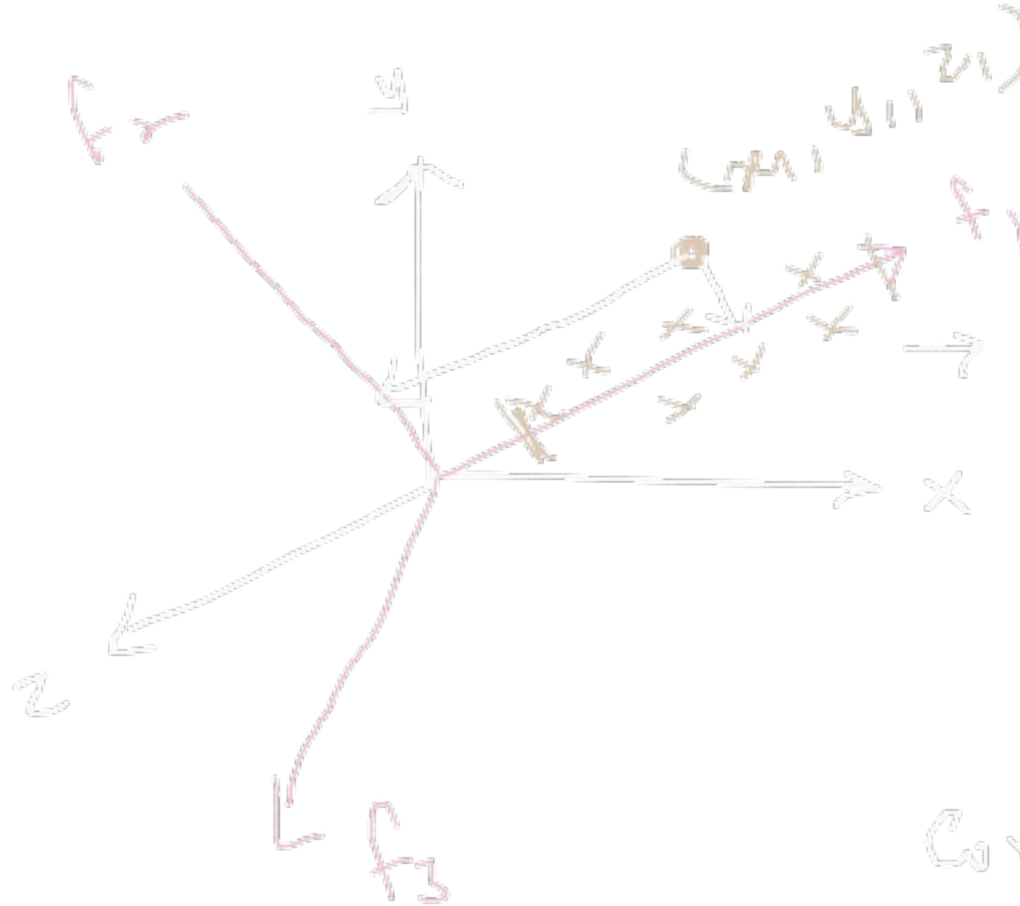
—	—	—
—	—	—
—	—	—
—	—	—

✓ ✓

85 10 5

PC1 PC2 PC3

—	—	—
—	—	—
—	—	—
—	—	—



$\rightarrow$ Find this direction

(1) Standardize the data ✓

Covariance Matrix ✓

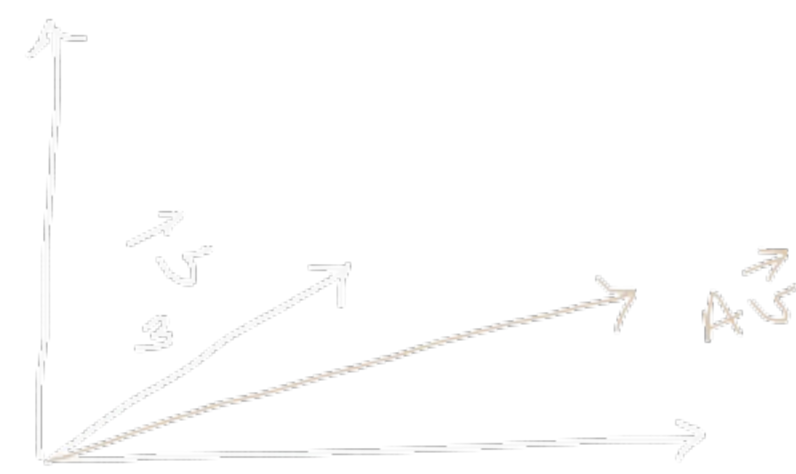
Correlation $\rightarrow$ -1 to $+1$

$$\underset{(x,y)}{\text{Correlation}} = \frac{\text{Covariance}(x,y)}{\sigma_x \sigma_y}$$

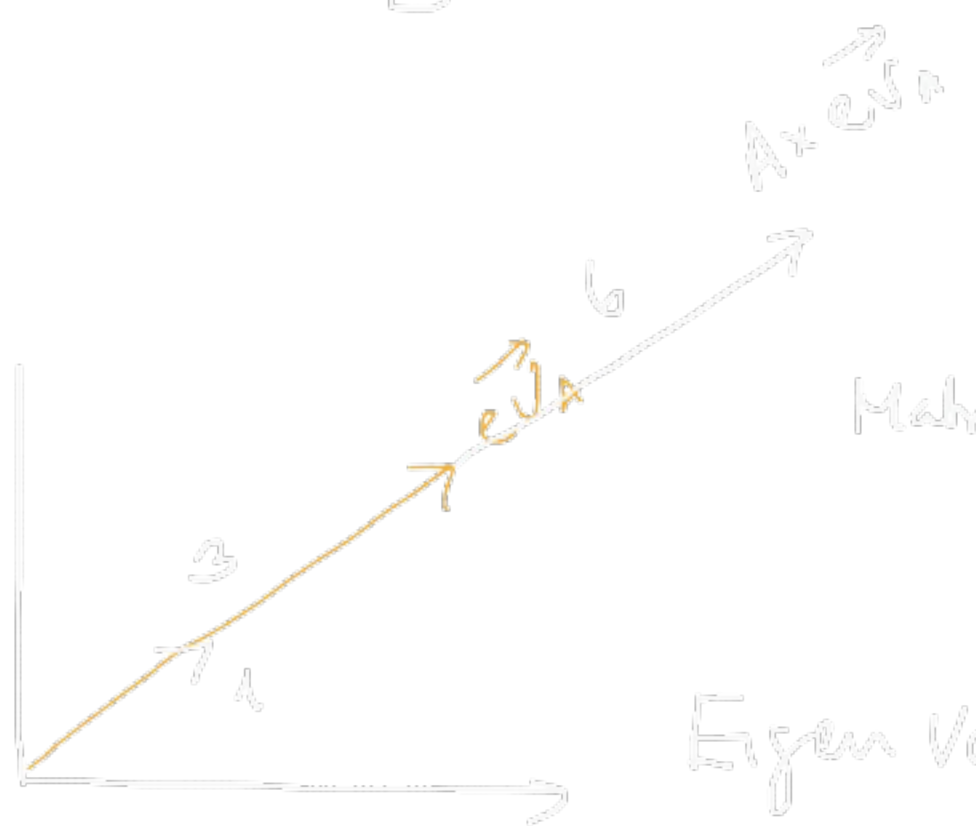
$$X = \begin{array}{c|c} & f_1 \; f_2 \; f_3 \; \cdots \; f_m \\ \begin{array}{c}1\\2\\3\\4\\ \vdots \\ n\end{array} & \left[\begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\right]_{n \times m} \end{array} \rightarrow \text{Cov}(X) \rightarrow \begin{array}{c|cccc} & f_1 & f_2 & f_3 & f_m \\ f_1 & V_1 & & & \\ f_2 & C_{21} & V_2 & & \\ f_3 & C_{31} & C_{32} & V_3 & \\ f_m & C_{m1} & C_{m2} & & V_m \end{array}_{m \times m}$$

Covariance $= +47$

$$
S = \begin{array}{c|c}
 & \begin{matrix} f_1 & f_2 & f_3 & \cdots & f_m \end{matrix} \\
\begin{matrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_m \end{matrix} &
\begin{bmatrix}
V_1 & C_{12} & & & \\
C_{21} & V_2 & & & \\
C_{31} & C_{32} & V_3 & & \\
 & & & \ddots & \\
C_{m1} & C_{m2} & & & V_m
\end{bmatrix}_{m \times m}
\end{array}
$$



$$A = \begin{bmatrix} 7 & 9 \\ 11 & 12 \end{bmatrix}_{2 \times 2}$$



$A \vec{v} = \lambda \vec{v}$

- A → Matrix
- λ → Scalar (eigen Value)
- $\vec{v}$ → Vector → eigen Vector

2 → Eigen Vals & Vectors

Eigen Vector of A → Vector which does not change direction, only dimension

$A_m \vec{v} = \lambda \vec{v}$

$S \rightarrow m$ eigen Values $(\lambda) \rightarrow$ Variances

m eigen Vector $(\vec{v}) \rightarrow$ direction of Variance

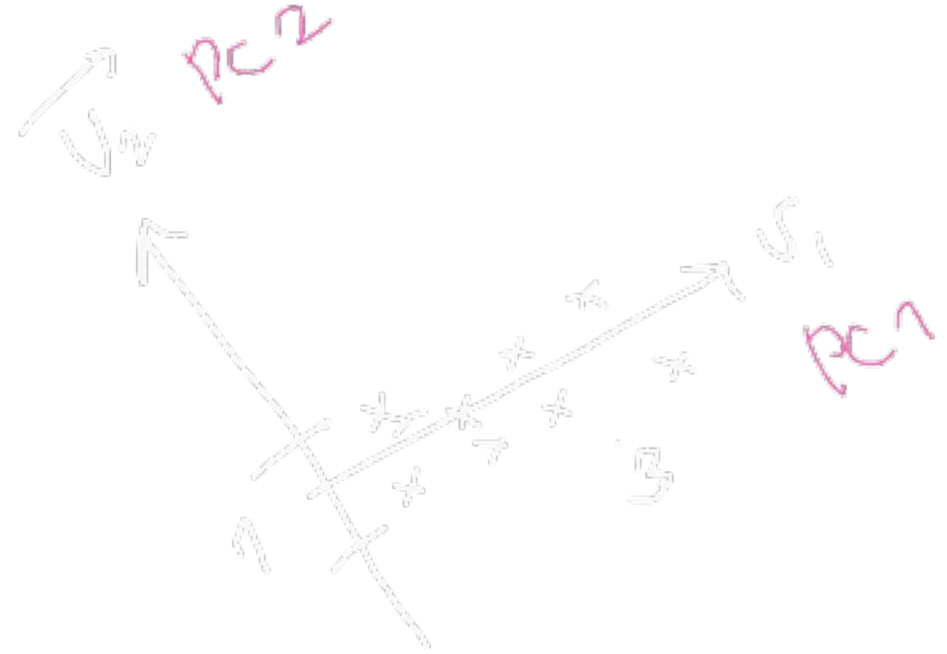
$\lambda_1 \quad \lambda_2 \quad \lambda_3$

$\vec{v_1} \quad \vec{v_2} \quad \vec{v_3}$

$\vec{v_1} \perp \vec{v_2} \perp \vec{v_3}$

$$\%\ \text{Info along } v_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \approx 100\%$$

$$\text{Info along } v_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2} \rightarrow 0\%$$



$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{3}{4} \rightarrow 75\%$$

$$\frac{\lambda_2}{\lambda_2 + \lambda_1} = 1/4 \rightarrow 25\%$$

				85%	10%	4%		1%
f_1	f_2	f_3	f_m	pc1	pc2	pc3		pcm

Eigen Value Decomposition

Step 1 Standardize the data

Step 2 Find the covariance matrix of X

Step 3 Decompose it into Eigen Values (m) & Eigen Vector (m)

Step 4 Find eigen Value which has max Value

The corresponding eigen

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

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

$$= \text{Variance}$$