$$\boxed{\max \frac{1}{n}\sum_{i=1}^{n} \vec{x}\cdot\vec{u}} \quad s.t. \ ||\vec{u}|| = 1$$

Constraint

Constrained optimization

* Unconstrained optimization Equ^n

$$\frac{1}{n}\sum_{i=1}^{n} \vec{x}\cdot\vec{u} + \lambda\left(||\vec{u}|| - 1\right)$$

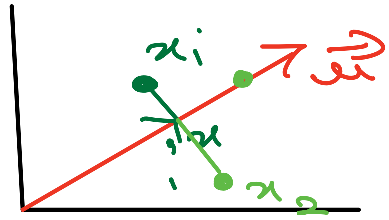
$\underbrace{||u|| = 1}$

lagrange's multiplier

* Gradient Ascent

Closed Form Solution

$$\max_{\vec{u},\lambda}\left(\frac{1}{n}\sum_{i=1}^{n} \underbrace{\vec{x}\cdot\vec{u}}_{-ve,\ +ve} + \lambda\left(||\vec{u}|| - 1\right)\right)$$



Can projection len be -ve?

$P_{x_2} \Rightarrow -ve$

$P_{x_1} \Rightarrow +ve$

$P \Rightarrow \frac{x_i \cdot \vec{u}}{\|\vec{u}\|}, \Rightarrow 1$

$\sum_{i=1}^{n} P_x \Rightarrow 0$

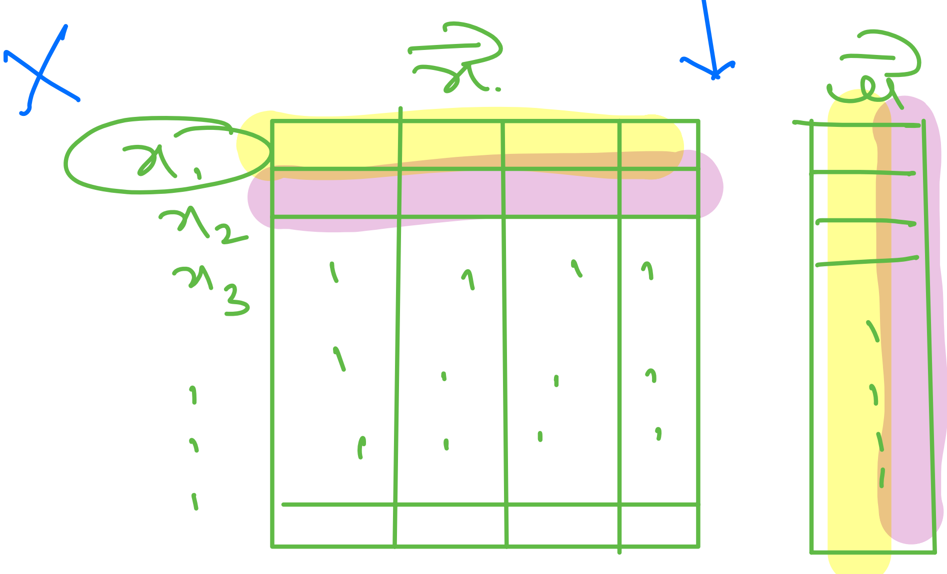
$\Rightarrow x_i \cdot \vec{u}$

-ve

+ve

Hence we can take Squares of Dot prod

$$\max_{\vec{u},\lambda} \left(\frac{1}{n} \sum_{i=1}^{n} (\vec{x_i} \cdot \vec{u})^2 + \lambda(\|\vec{u}\| - 1) \right)$$



$\Rightarrow$

$x_1 \cdot \vec{u}$

$+$

$x_2 + \vec{u}$

$x_3 + u$

$\vdots$

$x_n + \vec{u}$

$$\sum_{i=1}^{N} (\vec{x_i} \cdot \vec{u})^2 \Rightarrow (X \cdot \vec{u})^2$$

$\Rightarrow$ Hence we can remove Σ

$$\max_{\vec{u},\lambda} \left(\frac{1}{N} (X \cdot \vec{u})^2 + \lambda(\|\vec{u}\| - 1) \right)$$

$$(X \cdot \vec{u})^T (X \cdot \vec{u}) \Rightarrow A^2 \Rightarrow A^T \cdot A$$

$$u^T u$$

$$\max_{\vec{u},\lambda} \left(\frac{1}{N} (X \cdot \vec{u})^T (X \cdot \vec{u}) + \lambda(\vec{u}^T u - 1) \right)$$

$$(X^T \vec{u}^T) \times (X \vec{u})$$

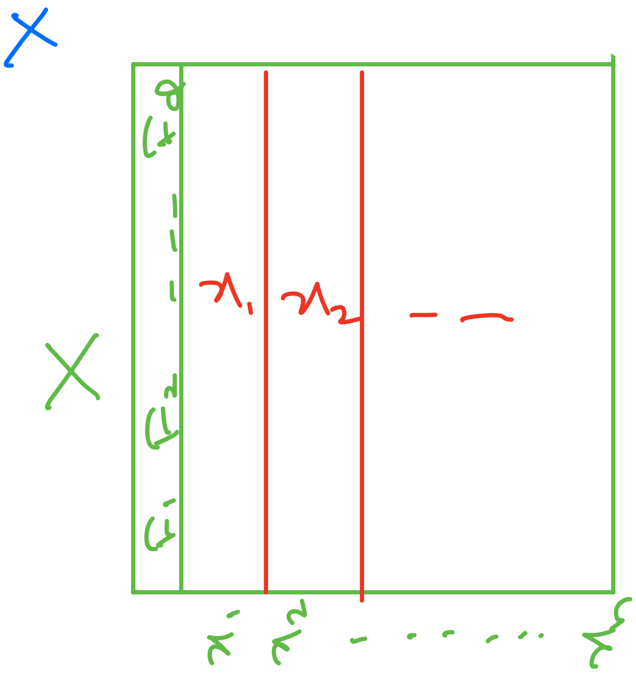
$$\max_{\vec{u},\lambda} \frac{1}{N} \left(\vec{u}^T \boxed{(X^T X)} \vec{u} + \lambda(u^T u - 1) \right)$$

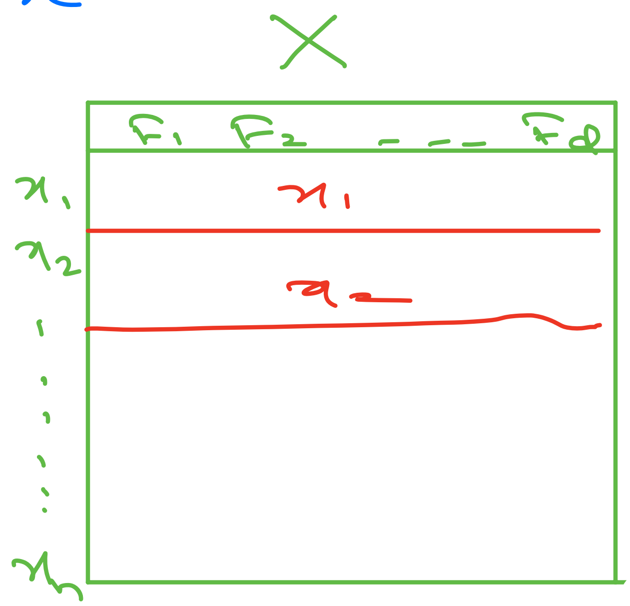
$$V \Rightarrow X^T X$$

$\arg\max_{w} \; w^T \left(\vec{w}^T V \vec{w} \right) + \lambda(\vec{w}^T\vec{w} - 1)$

Co-variance Matrix

$V_T \; x$





$$\Rightarrow \frac{x_1^2 + x_2^2 + x^3 \ldots x_n^2}{n}$$

$$\text{Cov} \Rightarrow \frac{(x_i - \bar{x}) \times (y_i - \bar{y})}{}$$

0 $\quad x_i \rightarrow 0$

$$\Rightarrow \frac{(x - \bar{x})^2}{n} \Rightarrow \left(\frac{x^2}{n}\right) \text{ Variance}$$

$= V \qquad = \text{input data } X$

$$(X^T \cdot X)/n$$

$$L = \boxed{\max_{\lambda,\, w} \quad w^T V w + \lambda (w^T w - 1)}$$

$$\frac{\partial L}{\partial \lambda} = \frac{\partial\, w^T V w}{\partial \lambda} + \frac{\partial \lambda (w^T w - 1)}{\partial \lambda}$$

$$= 0 + \frac{\partial (\lambda w^T w - \lambda)}{\partial \lambda}$$

$$= \frac{\partial (\lambda w^T w)}{\partial \lambda} - \frac{\partial \lambda}{\partial \lambda}$$

$$\frac{\partial L}{\partial \lambda} = w^T w - 1$$

$$\frac{\partial L}{\partial w} = \frac{\partial\, w^T V w}{\partial w} + \frac{\partial\, \lambda (w^T w - 1)}{\partial w}$$

$$= \frac{\partial\, \lambda w^T w}{\partial w} - \frac{\partial \lambda}{\partial w} \rightarrow 0$$

$$\boxed{\frac{\partial L}{\partial \vec{u}} \Rightarrow 2uV + 2\lambda u}$$

Setting the partial Derivatives to 0

$$\frac{\partial L}{\partial \lambda} \Rightarrow u^T u - 1 \Rightarrow 0$$

$$u^T u = 1 \qquad \|u\| \Rightarrow 1$$

$$\boxed{\frac{\partial L}{\partial \vec{u}} \Rightarrow 2uV + 2\lambda u} = 0$$

$$2\vec{u}V + 2\lambda\vec{u} = 0$$

$$\cancel{2}\vec{u}V = -\cancel{2}\lambda\vec{u}$$

$$\underline{\vec{u}V = -\lambda\vec{u}}$$

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

Dataset matrix

Vector

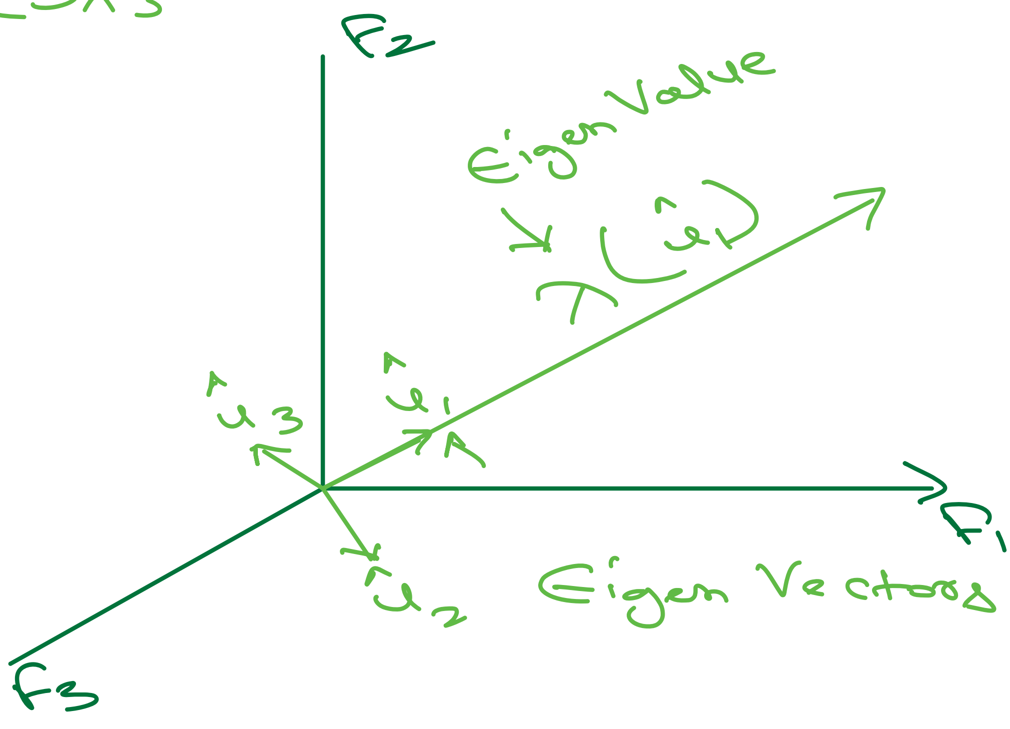
Constant Value

need to find

V_{matrix} (3×3)

λ's — 3 λ's

$\hat{u}$'s — (3 $\hat{u}$'s



How can we solve

$X \Rightarrow 100, 3 \Rightarrow 100 \times 3$
(rows) (d)

$V \Rightarrow X^T X \Rightarrow \underbrace{(3 \times 100) \times (100 \times 3)}$

$V \Rightarrow 3 \times 3$ $\begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \end{bmatrix}$

$\Rightarrow$ np.matmul(X.T, X)/N $\Rightarrow$
or
$\Rightarrow$ np.cov(X)

$V_{3\times3}$ — $\lambda_1, \lambda_2, \lambda_3$ (EigenVals)
— $\vec{u}_1, \vec{u}_2, \vec{u}_3$ (EigenVectors) (PC's)

$\lambda_1 \rightarrow \vec{u}_1$, $\lambda_2 \rightarrow \vec{u}_2$, $\lambda_3 \rightarrow \vec{u}_3$

λ's of each Eigen Vector means the corresponding importance

Ex:
$\Rightarrow$ $\vec{u}_1, \vec{u}_2, \vec{u}_3$
$\lambda_1 = 2 \quad \lambda_2 = 5 \quad \lambda_3 = 0$

$\vec{u}_2$ is most important

Information associated with $\vec{u}$'s

X ⇒ 5 features ⇒

λ	$\vec{u}$
10	$\vec{u}_1$ ⇒ [0.5, 0.2, ...]
7	$\vec{u}_2$
6	$\vec{u}_3$
3	$\vec{u}_4$
1	$\vec{u}_5$

$\vec{u}_2 \ldots \vec{u}_5$ ⇒ PC's

m → m P.C.A

'information / Variance Stored' in

$$\vec{u}_i \Rightarrow \frac{\lambda_i}{\sum_{i=1}^{d} \lambda_i} \Rightarrow \text{Explained Variance Ratio}$$

$$\frac{\lambda_i}{\sum_{i=1}^{d} \lambda_i} \Rightarrow 100\%$$

$$\vec{u}_2 \Rightarrow \frac{7}{10+7+6+3+1} \Rightarrow 0.25$$

⇒ 25% 'information / Variance' is Stored 'in U_2'

$$\vec{u}_1 \Rightarrow \frac{10}{27} \Rightarrow 37\%$$

Q ⇒ 5

i) if keep only first two PC's
How much info will i preserve

$\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4, \vec{u}_5$

remove remaining

$$\frac{\lambda_1}{\sum_{i=1}^{n} \lambda_i} + \frac{\lambda_2}{\sum_{i=1}^{n} \lambda_i} \quad 0.37 + 0.25$$

0.62

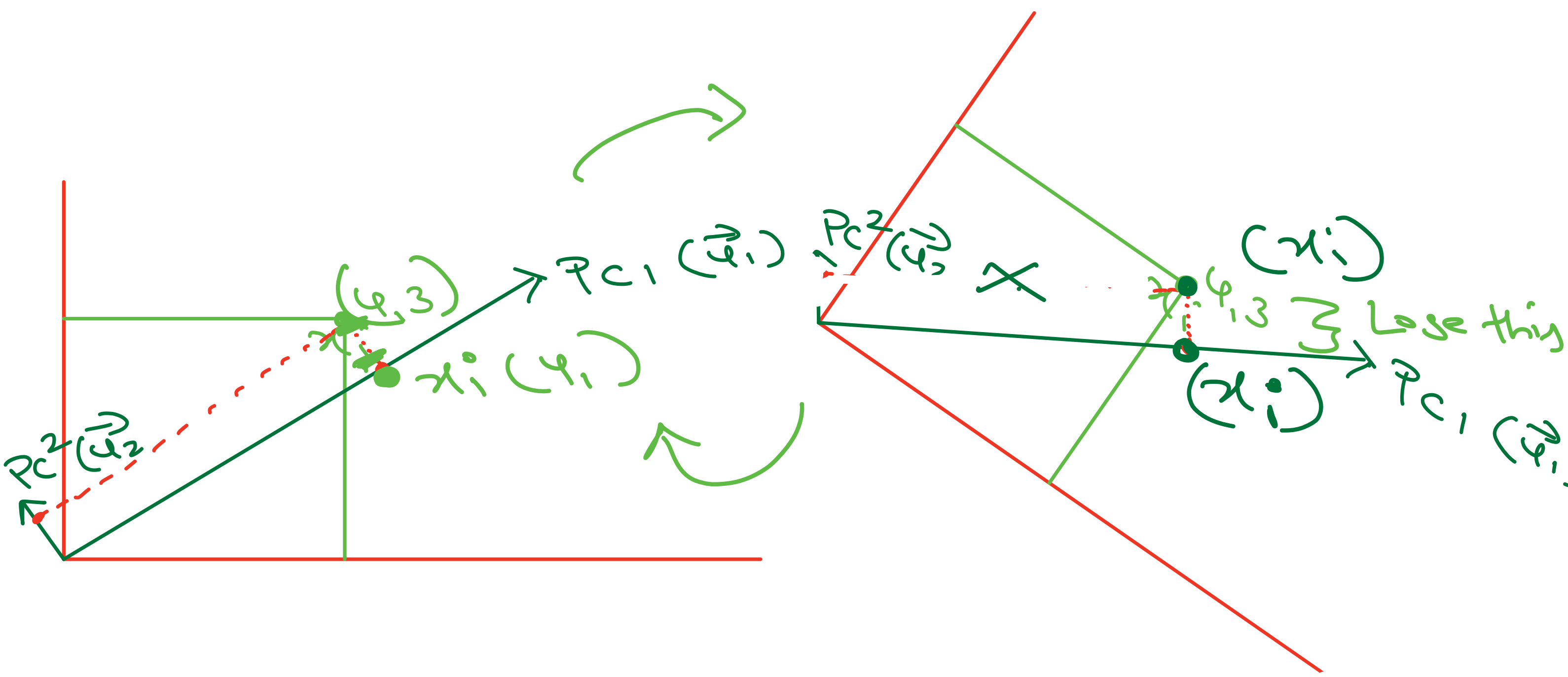
62%

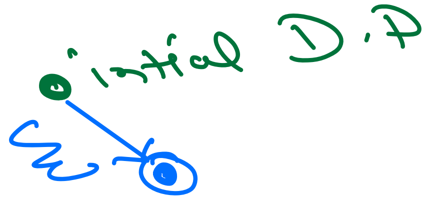
2 pc

$100 - 62 = 38\%$

ii) 90% information ⟶ How many min PC's will I need

$\boxed{\text{Information Loss}}$





Summary:

- There is Trade off between
    - Dropping PC's to reduce
    - Information Loss

In Next Session

- Issues with PCA
- How T-SNE is different from PCA
- U-MAP

- Maths for ML
  (Vectors)
  ↳ first 5

- Dot product
- Projection of Vector
- Norm of Vector