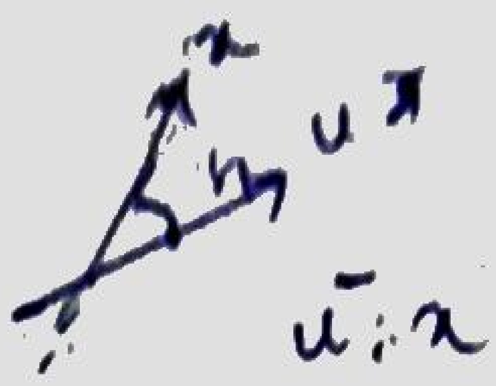
• PCA, we find the direction of maximum variance



$u \cdot x = u^T x$

$\bar{z} = \frac{1}{N} \sum_i z_i$

$= \frac{1}{N} \sum_{i=0}^{N} u^T x_i$

$\bar{z} = u^T \frac{1}{N} \sum_{i=0}^{N} x_i$

$\bar{z} = u^T \mu$

$z = u^T x_i$

1xd, dx1

$z = \mu^T u$

1x1 = 1xd x dxk

• $\underset{u}{\text{Maximize}} \; \frac{1}{N} \sum_{i=1}^{N} (z_i - \bar{z})^2$

$= \frac{1}{N} \sum_{i=1}^{N} (u^T x_i - u^T \mu)^2$

$= \frac{1}{N} \sum_{i=1}^{N} (u^T (x_i - \mu))^2 \rightarrow \frac{1}{N} \sum_{i=1}^{N} (u^T x_i - \mu^T \mu)(x_i^T u - \mu^T \mu)$

$\frac{1}{N} \sum_{i=1}^{N} u^T [x_i - \mu][x_i - \mu]^T u$ (dx1, x1xd)

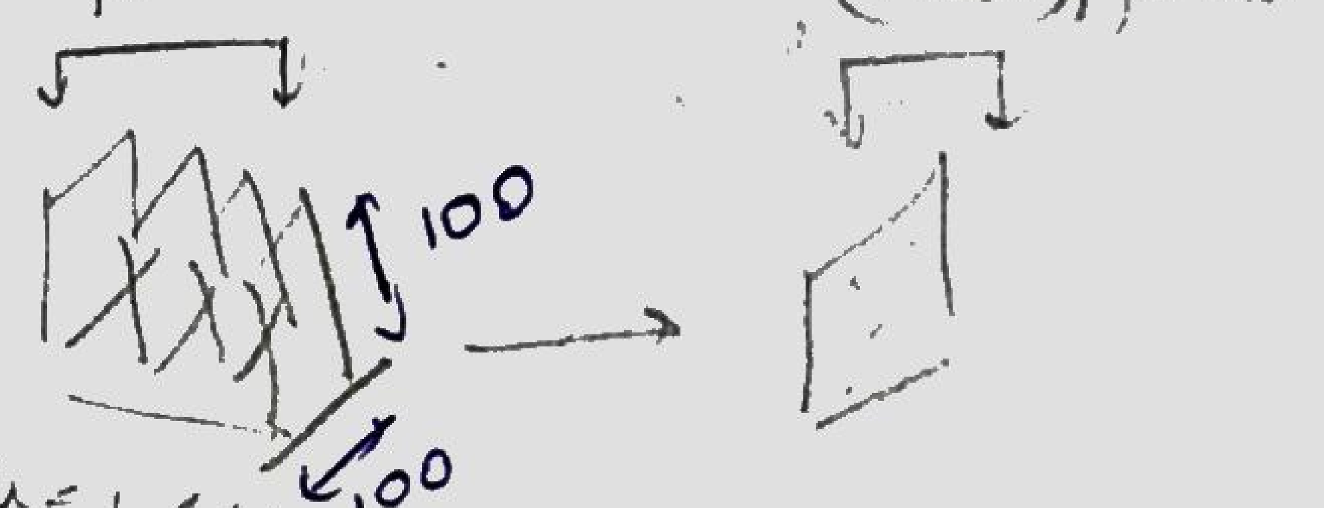
$\frac{1}{N} \sum_{i=1}^{N} u^T \Sigma [x_{ij} - \mu_j](x_{ij} - \mu_j)^T$

$= u^T \Sigma u$, such that $u^T u = 1$

↓ via lagrangian method,

u is the eigenvector of the largest eigen value λ.

Eigen-face:-



$\mu = \frac{1}{N} \sum x_i$

Each feature is a photo ⇒ d = No. of points (pixels) $= 10^2 \times 10^2 = 10^4$

$\dim(\Sigma) = d \times d = 10^4 \times 10^4$

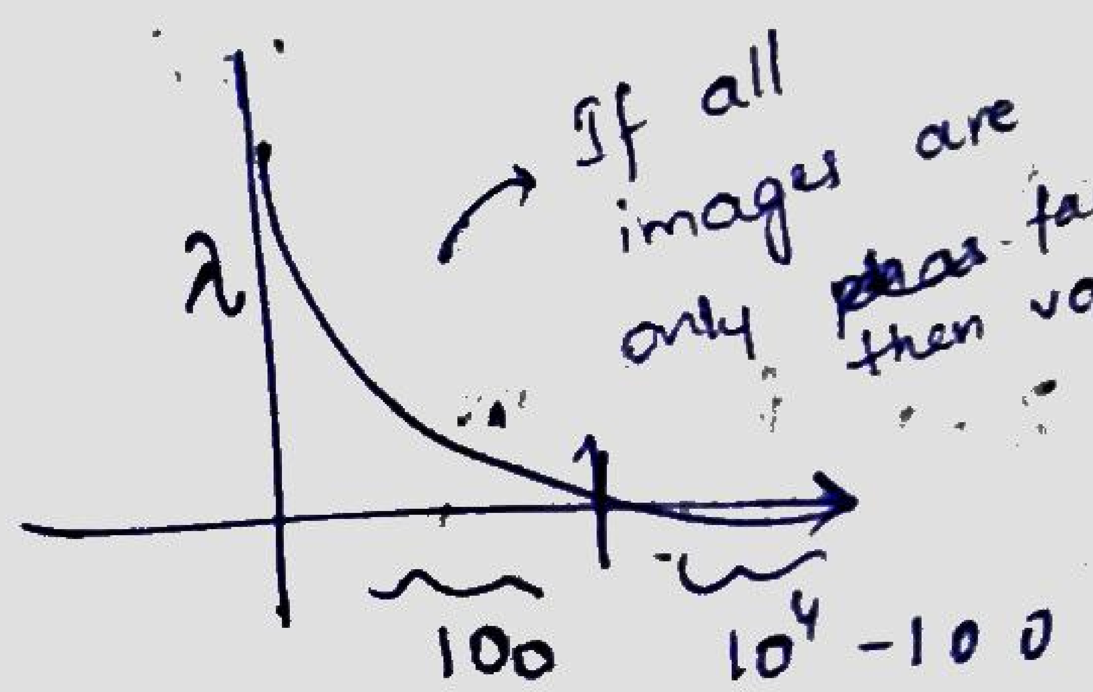
Finding eigen vectors of $10^4 \times 10^4$ matrix is difficult.

No. of photos = 1000

Max vectors = 10^4

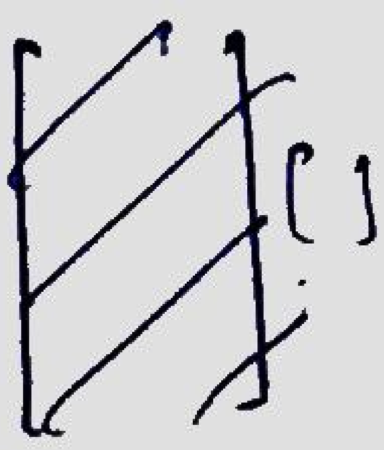
But as almost all photos are same,

a few eigen vectors



$u = [10^4 \times 100]$ (k)

$z = \begin{bmatrix} x^T u_1 \\ x^T u_2 \\ \vdots \\ x^T u_{100} \end{bmatrix}$ →



$z = x^T u$

$\dim(10^4 \times 1)$

So $\dim(z) = 100 \times 1$

• $x = \sum_{i=N}^{10000} \alpha_i u_i \rightarrow$ eigen face.

$\downarrow$ A

protection $x^T u_i$

New face $= \sum_{i=N}^{10000} \alpha_i u_i + \sum_{i=\text{rest}} \alpha_i u_i \Longrightarrow$ Useless

Almost similar for every photo

• Practical eg: A video conference: If only the 100 numbers are sent, your ~~features~~ face can be drawn, but why don't we use?

• This method is valid iff the images we capture only faces and not other objects like cars, ---

• So not useful

In the photo example, $10^4 > 10^3$
$d > N$.

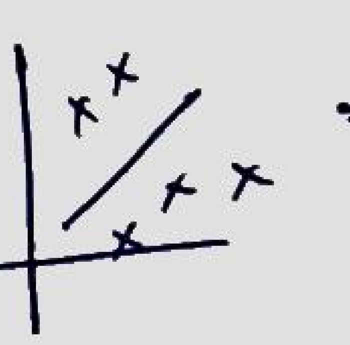
Think of XX^T and X^TX

$X = N \times d$, $X^T = d \times N$

$XX^T = N \cdot N$

$X^TX = d \cdot d$

• $||^{ly}$ decay affect for line



$\Rightarrow \lambda_2$ need be small

• If identical faces $\Rightarrow$ steeper curve

$XX^T u = \lambda_1 u$

$X^TX \, v = \lambda_2 u$

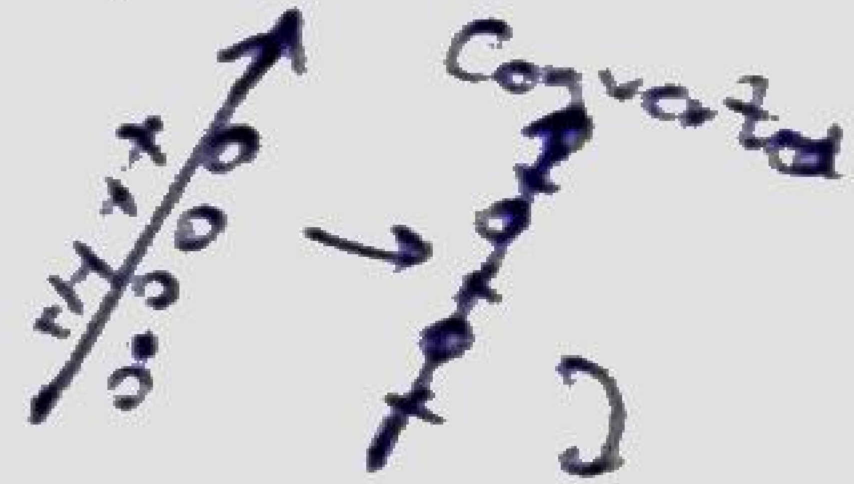
$X^TXX^Tu = \lambda_1 X^Tu$

$X^TX \, v = \lambda_1 v$

$\Rightarrow$ Compute eugen vectors of $N \times N$ matrix $= 10^3$ and convert $d \times d \Rightarrow 10^4$

Image- is Positive Semi-Definite

Classification + Dimensionality Reduction :-



If  → 

Perpendicular

Don't PRESUME → (But ⊥lar need not always be best soln.)

Miserably failed

Unsupervised → You don't care, which eigen vector was used.

Supervised →

LDA| Fisher.

Where class info is used

10/9/18

Lec-11

- (x_i, y_i) $i = 1..N$

$x_i \in R^d$

$\boxed{Q: \text{Is } N > d \text{ or } d > N}$ $\Rightarrow$ If d is large, in the starting, need not be the best or right x.

- A need for feature Selection/ Extraction:-

Subset Selection

- Can you find a subset of selections, that can do almost as good as the original set.
- To reduce cost.
  - ↳ Example he took is, doing medical tests to get info abt patient
- Subset selection is a hard prob and so, finding the best subset would be hard

Extraction

$z = Ux$ [A unitary transformation]

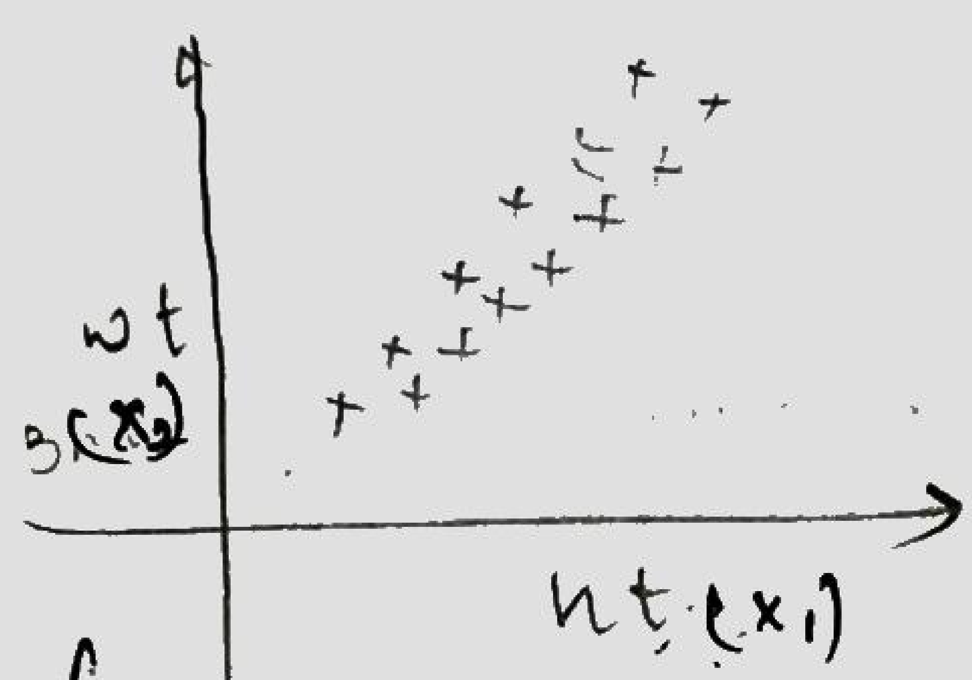
$k < d$

- Sort, of like combining medical tests, to ~~get~~ ~~a~~ reduce the ~~nott~~ dimensions
- // A Dimensionality Reduction
- ~ Can be thought as extracting new features for the task
- Sensing cost → no change
  But downstream → Less of feature extraction

=

⇓

$\boxed{\text{Soln: PCA}}$



- Problem:-
  - ↳ ~~Correla~~ Intelligence vs Ht & wt
  - ↳ Suppose this Ht & wt of a single child & ~~plotting~~
  - ↳ We observe H, W are linearly correlated

$R^{d \times n}$ of dimensions

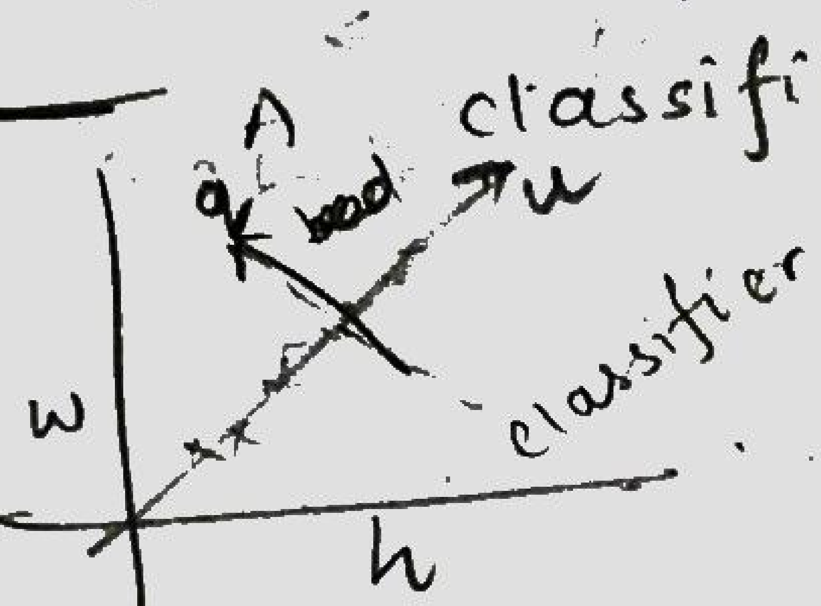
———×××××××××→ (z_1)

age

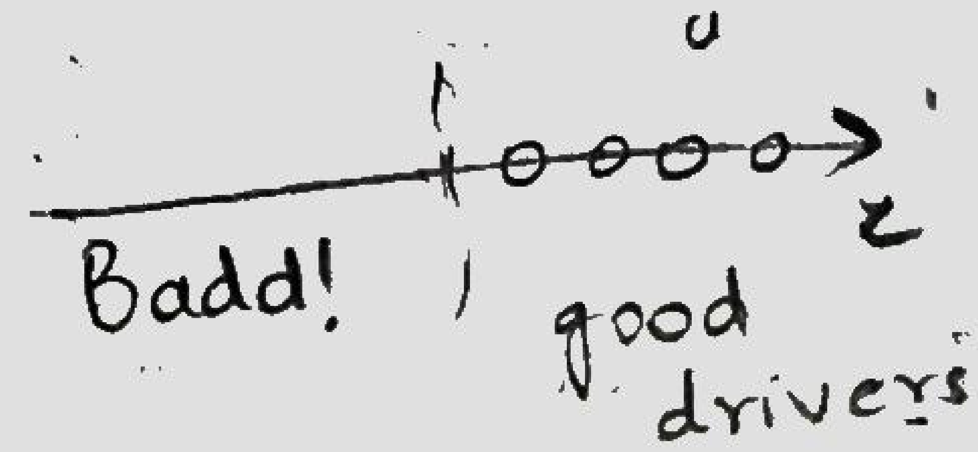
comb of wt + ht

$z \to$ is practically capturing ~~x₁~~ and x_2

classifier → Capability of driving



- - ->



If projected in q(bad) direction

No age separ