INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD
OUTREACH DIVISION
talentsprint

Special Lecture

Dimensionality Reduction, PCA, EigenFaces

Higher Dimensional Visualisation

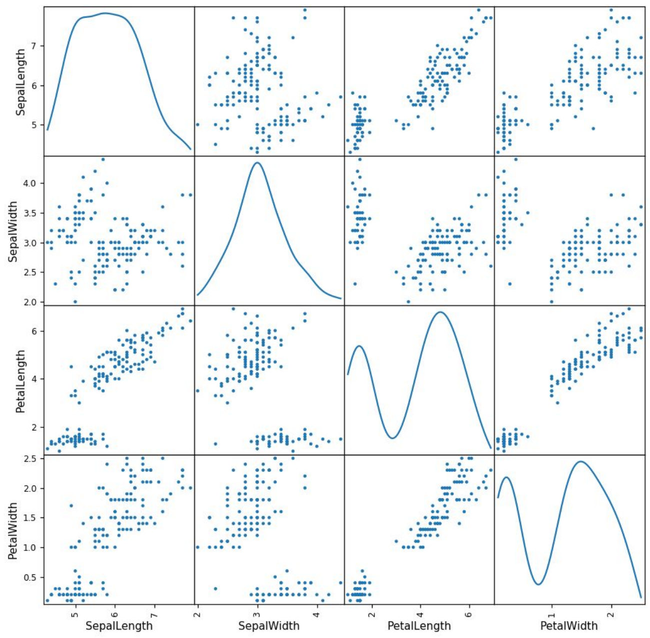
Visualisation

- Select effective *visual encodings* to map data values to graphical features
- Features such as position, size, shape, and color.
- Challenge :
    - for any given data set the number of visual encodings is extremely large

Higher Dimensional Visualization Techniques

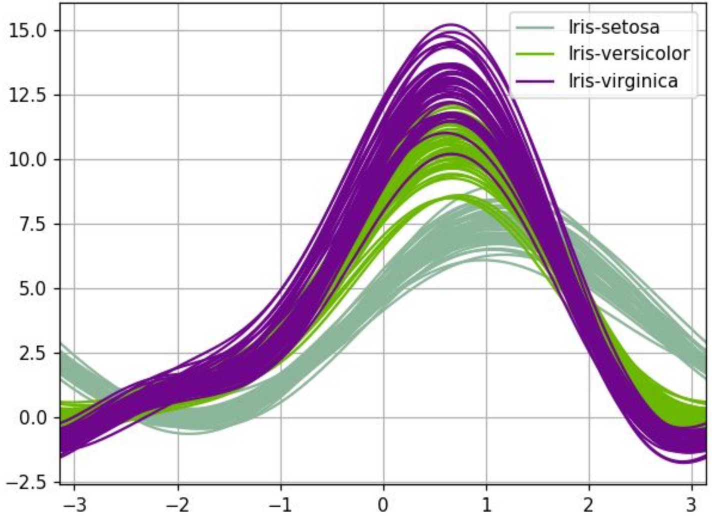
Scatter Plot

- A Scatter plot displays the correlation between a pair of variables
- These scatter plots ($^{n}C_{2}$ pairs) can be organized into a matrix, making it easy to look at all pairwise correlations in one place



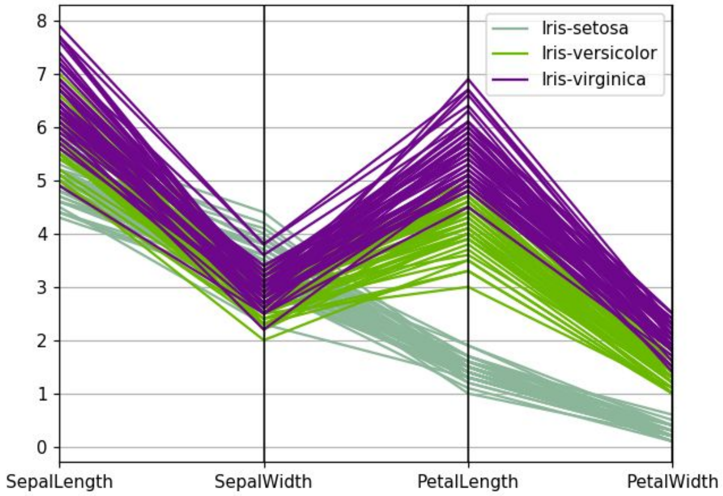
Andrew Plot

- An Andrews plot is a graphical data analysis technique for plotting multivariate data.
- An Andrews curve applies the following transformation to a set of data
    - **$F(t) = X_1/sqrt(2) + X_2 sin(t)+X_3 cos(t)+X_4 sin(2t)+X_5 cos(2t)+....$**
    - **Where $-\Pi<t<\Pi$**



Parallel coordinates Plot

- Parallel coordinates is a visualization technique used to plot individual data elements across many dimensions
- Each of the dimensions corresponds to a vertical axis
- Each data element is displayed as a series of connected points along the dimensions/axes



Reducing Dimensions to visualize

- Feature Selection :
  - Choose the "best" features from your data, which you then visualize
- Feature Extraction :
  - Initial set of measured data and builds derived features intended to be informative and non-redundant, facilitating

Selecting Features as Matrix Multiplication

- Selecting first and third feature

$$\begin{bmatrix} x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

- Selecting first and fourth feature

$$\begin{bmatrix} x_1 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$X' = AX$$

Selecting Features as Matrix Multiplication

- A "new" set of features are selected/extracted from original one by a matrix multiplication.
- Rows of **A** decide what the new features are. (They need not be 0 and 1)
- Often rows of **A** is smaller than column of **A**. This is also called **dimensionality reduction**.

Feature Selection and Extraction

- Selection:
  - Select some features out of a pool
    - (Simple A with 0/1)
- Extraction:
  - Extract a set of new features
    - (elements of A need not be 0/1)

Feature Selection and Extraction

Extraction is often required:

- To visualize in 2D/3D.
- To remove some “useless” or “less useful” features.
- Make computations efficient

(Note: original data could be 1000s of dimension!!)

Principal Component Analysis

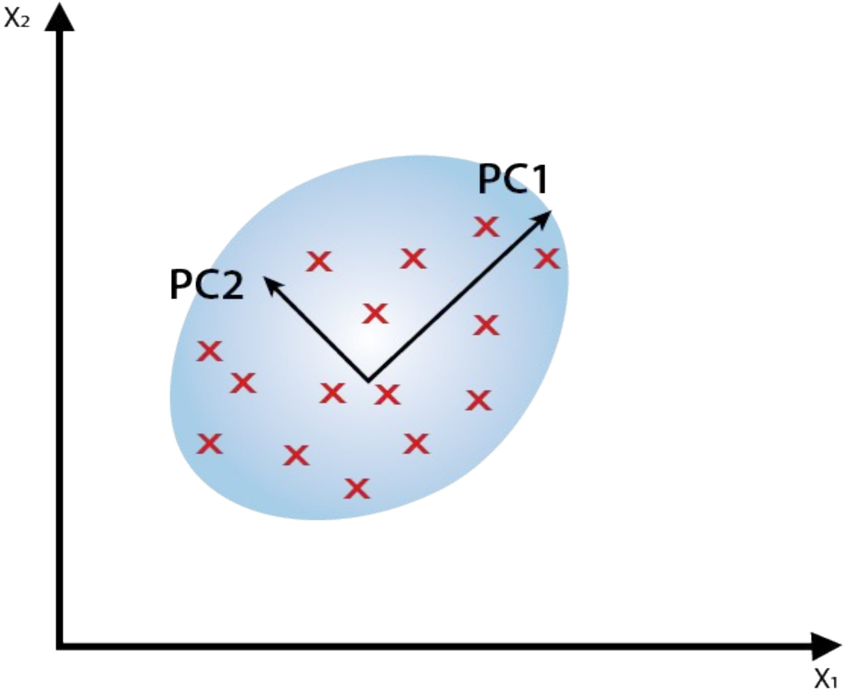
Simplifying Representations

Outline

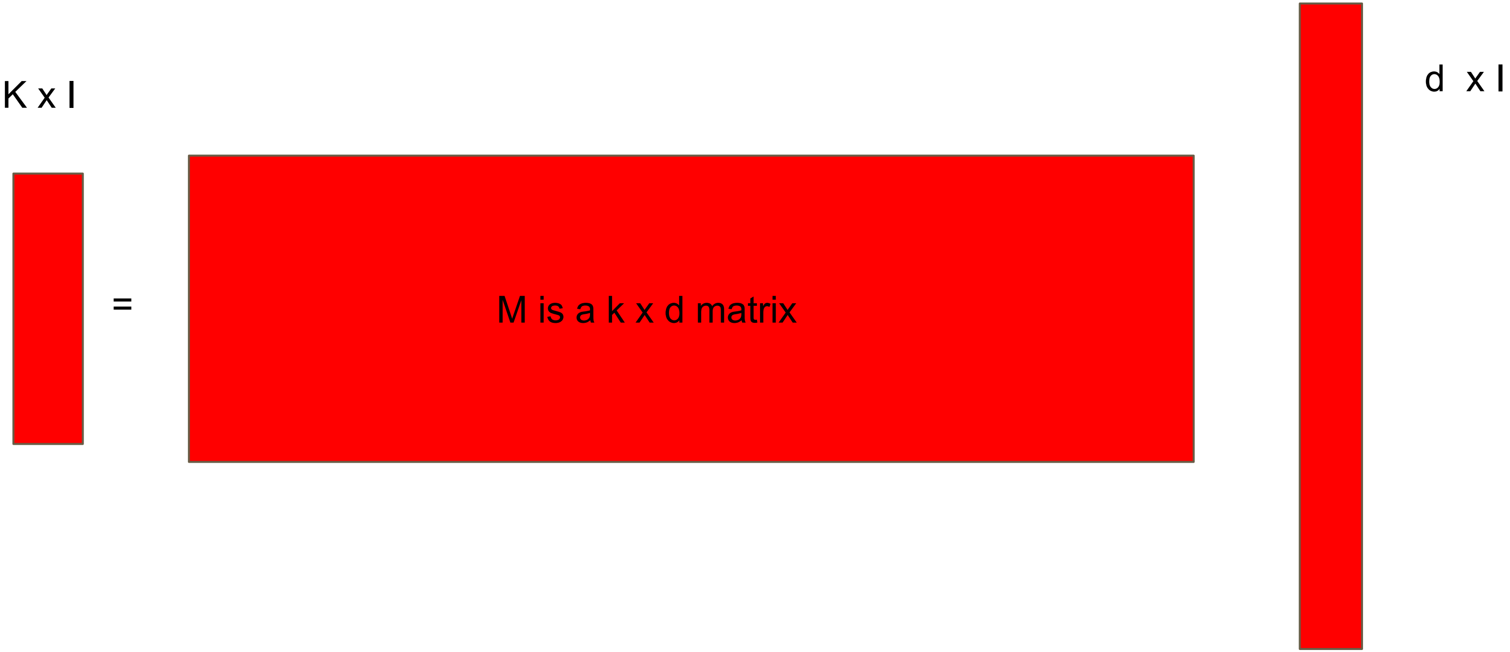
- What is Principal Component Analysis
- PCA based feature extraction
- Projection of point to a line
- PCA and Covariance Matrix

Principal Component Analysis

- PCA is used to reduce the dimensionality of data



PCA based Feature Extraction



PCA based Feature Extraction

- Let $\mathbf{X_i}$ be an image represented as column vector

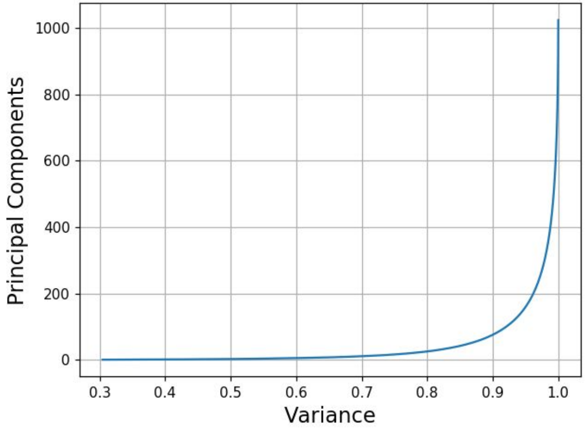
- Let $\mathbf{X} = [x_1 , x_2 , ..., x_N]$ be a zero mean matrix of size **d X N**

- Let $A = XX^T$ be a ***d X d*** Matrix. It also has ***d*** edge vectors each of ***d*** dimensions.

- Row of matrix "**M**" are the selected **K** Eigenvectors of the matrix **A**.

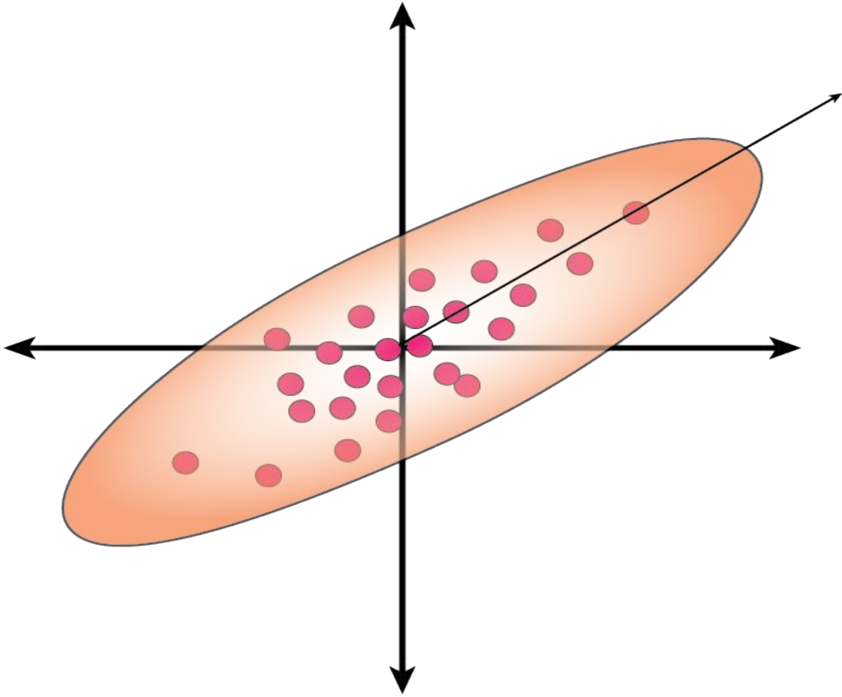
PCA based Feature Extraction

- Plot Principal components vs Variance retained
- Usually a small number of Principal components are enough to retain the variance considerably
- Most of them are near zero or even zero
    - Eg: K = 500 r = 0.9



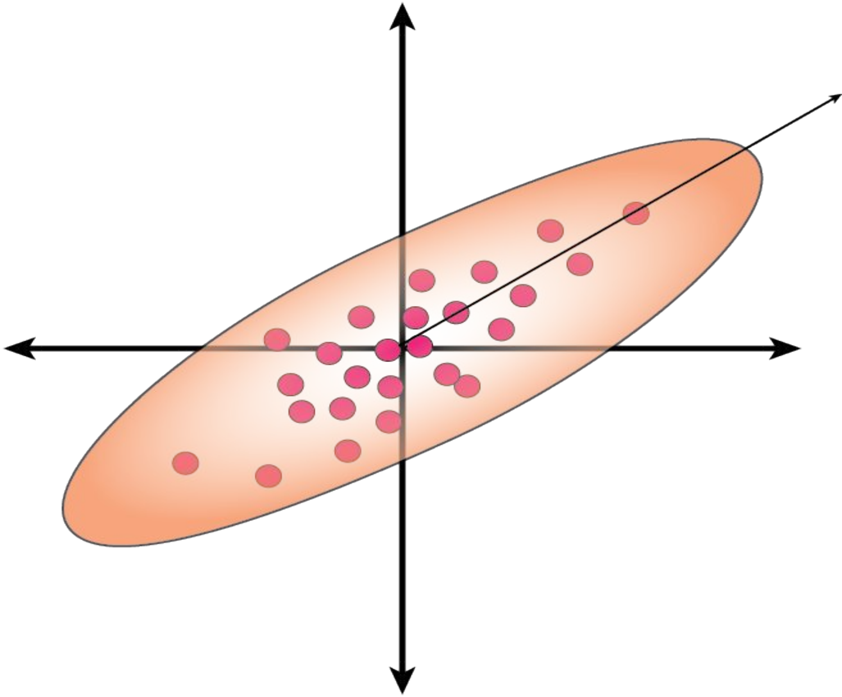
Projection of Point to a Line

- Consider a line through origin
  - Vector along the line
- What does dot product mean?
- Dot product with which vector?



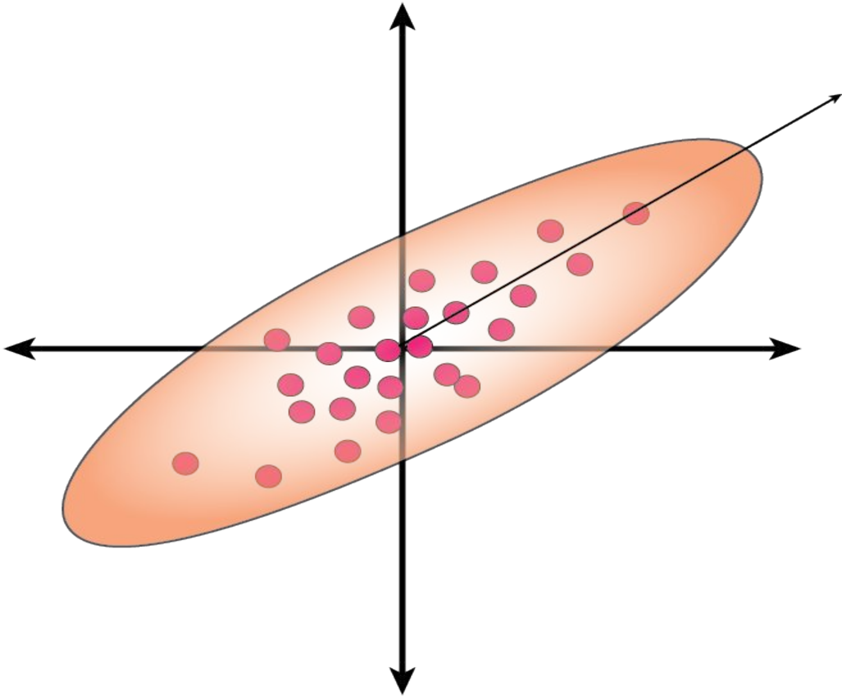
Projection of Point to a Line

- Vector along the line or line coefficients?
  - We consider dot product with vector along the line here
  - We talked about dot product with coefficients



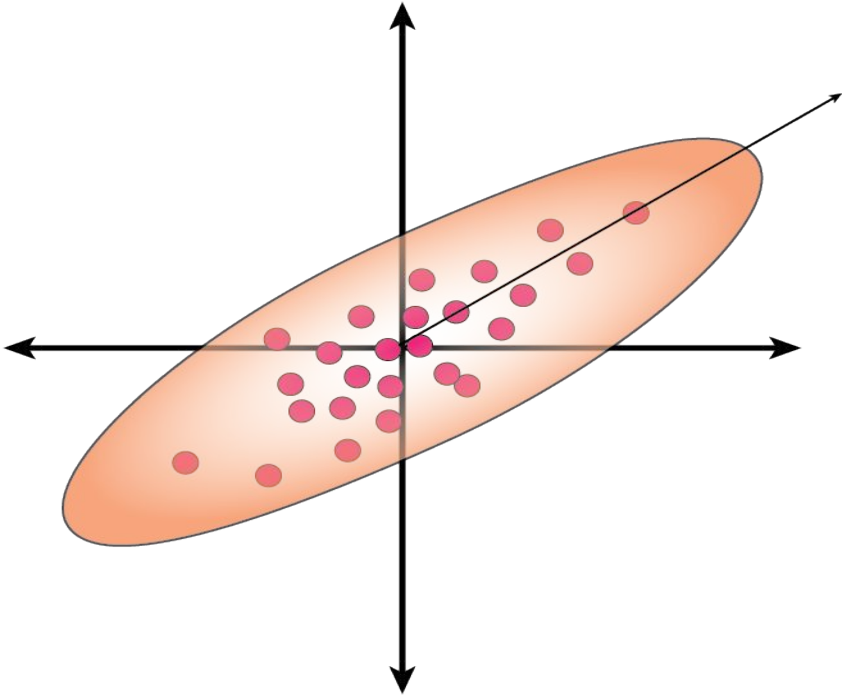
PCA and Covariance Matrix

- Going to 2D to 1D
- Which feature to select ?
  - This may be any feature(or vector in any direction)



PCA and Covariance Matrix

- Two viewpoints:
  - Maximize variance
  - Minimize error



PCA and Covariance Matrix

- Solution to both happens to be :

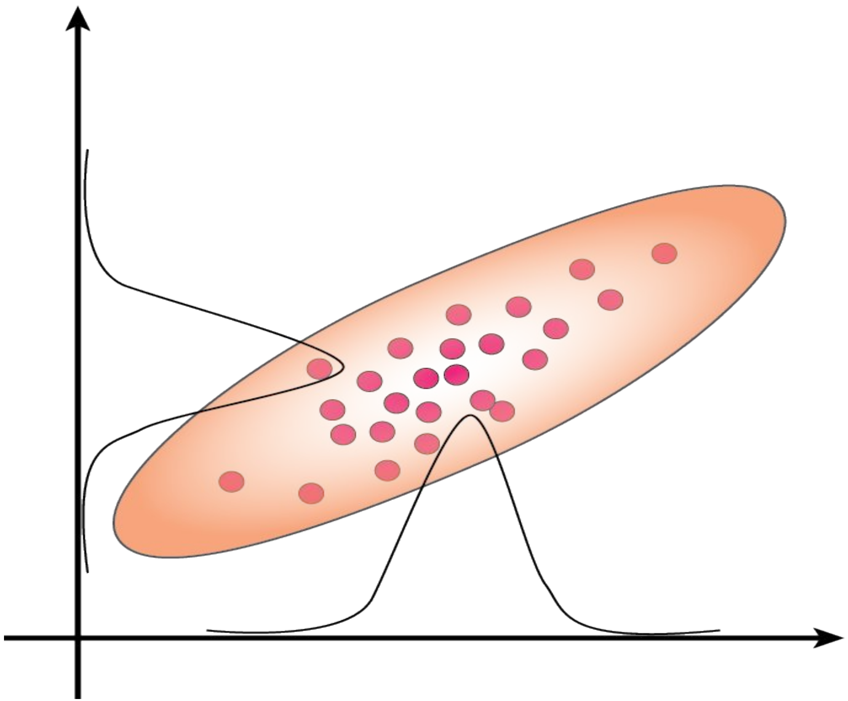
$$Ax = \lambda x$$

- What is A? How do we solve this?

PCA and Covariance Matrix

- We have variances along each dimension

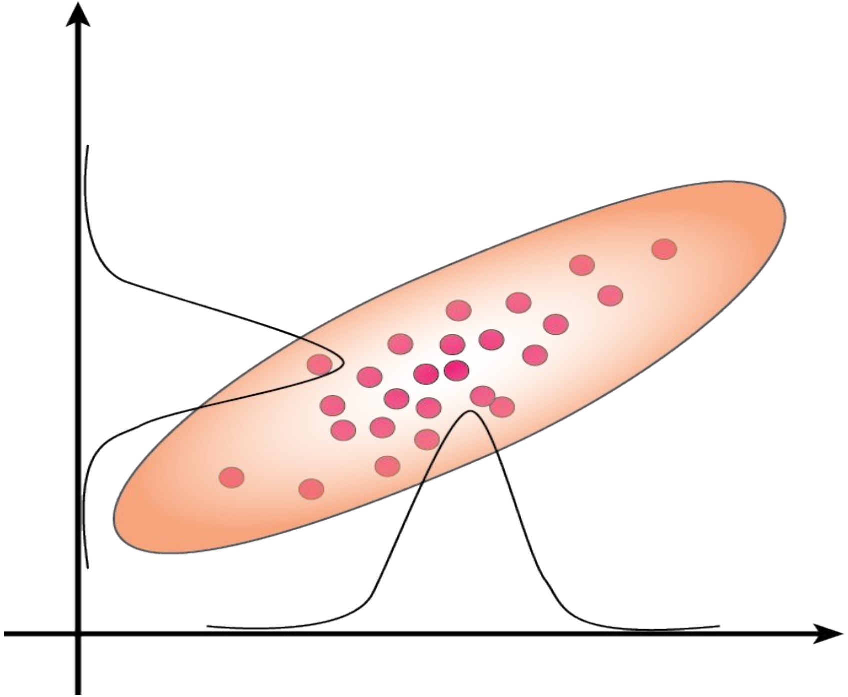
- The samples also co-vary. i.e, features are not independent



PCA and Covariance Matrix

- Captured using a covariance matrix

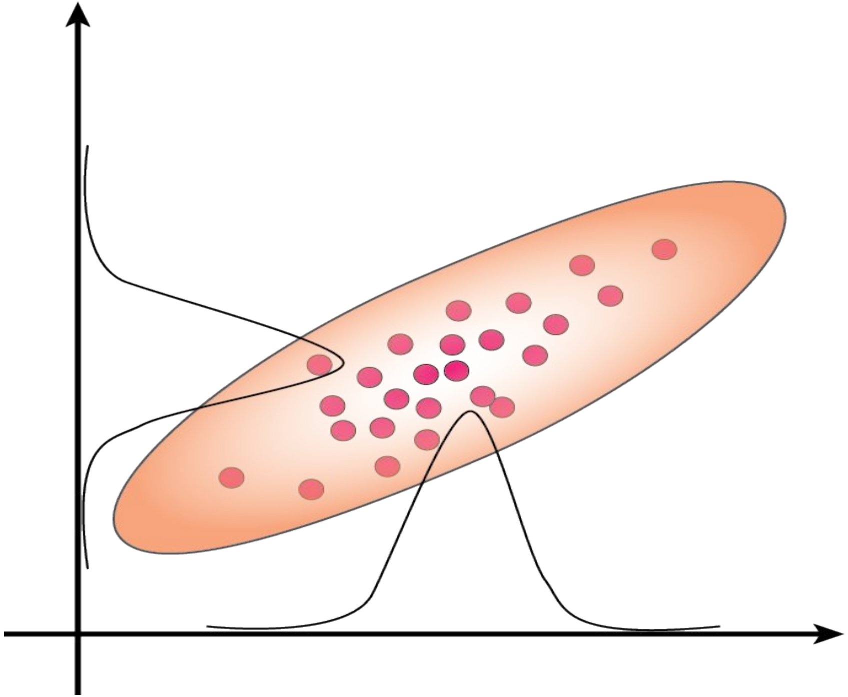
$$
\begin{bmatrix}
V_a & C_{a,b} & C_{a,c} & C_{a,d} & C_{a,e} \\
C_{a,b} & V_b & C_{b,c} & C_{b,d} & C_{b,e} \\
C_{a,c} & C_{b,c} & V_c & C_{c,d} & C_{c,e} \\
C_{a,d} & C_{b,d} & C_{c,d} & V_d & C_{d,e} \\
C_{a,e} & C_{b,e} & C_{c,e} & C_{d,e} & V_e
\end{bmatrix}
$$



PCA and Covariance Matrix

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

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



Likelihood Function

$$N\left(x,\mu,\sigma\right)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Becomes

$$N\left(x,\mu,\sum\right)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\sum\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\sum^{-1}(x-\mu)}$$

Challenge of Data

$$\begin{bmatrix} V_a & C_{a,b} & C_{a,c} & C_{a,d} & C_{a,e} \\ C_{a,b} & V_b & C_{b,c} & C_{b,d} & C_{b,e} \\ C_{a,c} & C_{b,c} & V_c & C_{c,d} & C_{c,e} \\ C_{a,d} & C_{b,d} & C_{c,d} & V_d & C_{d,e} \\ C_{a,e} & C_{b,e} & C_{c,e} & C_{d,e} & V_e \end{bmatrix}$$

- 1-dim had 2 parameters to estimate
- d-dim will have not just 2d, but over $d^2/2$ parameters.

Thanks!

Questions?

EigenFaces

Face Recognition

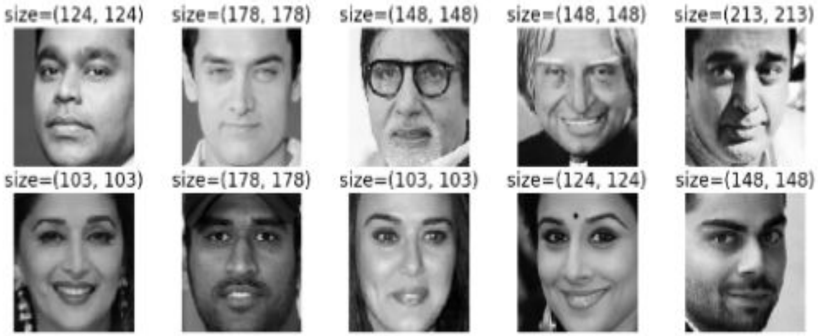
Problem Statement

Step 0: Preprocess to

224 X 224 or 50716 X 1

Recognize Indian celebrities

- 10 Classes
- 20 Example each(15 for training and 5 for testing;No colour)



Representations

- Eigen faces(classic)
- VGG Deep Net Features(Modern)

Eigen Face : Visualisation

Mean of all faces

Top 20 EigenVectors(EigenFaces)

EigenFace Feature

X

- Any face in the database can be accurately represented as a linear combination of these 20 eigen faces.

$$x = a_1 e_1 + a_2 e_2 .... a_{20} e_{20}$$

Thanks!

Questions?