CS6640 Week 10 — 29 Oct '20

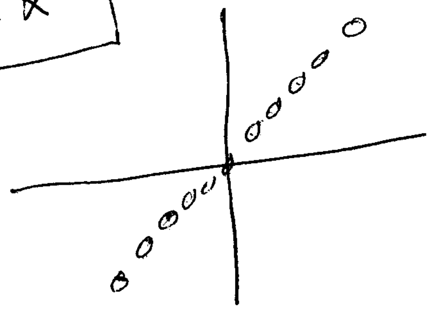
Principal Component Analysis

Find lowest dimensional space which "adequately" represents the given data (as vectors)

Consider pts:

$$\begin{bmatrix} -5 & -5 \\ -4 & -4 \\ \vdots & \vdots \\ 4 & 4 \\ 5 & 5 \end{bmatrix}$$

$\boxed{y = x}$



Then the covariance of these pts is

$$C = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{bmatrix}$$

where $K_{x_i x_j} = \operatorname{cov}(x_i, x_j) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_i)(x_j - \bar{x}_j)$

$\bar{x}_i$ is mean of x_i and $\bar{x}_j$ is mean of x_j

for pts:

$$C = \begin{bmatrix} 11 & 11 \\ 11 & 11 \end{bmatrix}$$

$$= \begin{bmatrix} 11x + 11y \\ 11x + 11y \end{bmatrix} = \begin{bmatrix} 11 & 11 \\ 11 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Usually need to transform pts to have 0 mean

E.g., consider 2D set of points

$$pts = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \\ x_n & y_n \end{bmatrix}$$

$\bar{x}$ mean of x

$\bar{y}$ mean of y

$$pts0 = \begin{bmatrix} x_1-\bar{x} & y_1-\bar{y} \\ x_2-\bar{x} & y_2-\bar{y} \\ \vdots & \vdots \\ x_n-\bar{x} & y_n-\bar{y} \end{bmatrix} = \begin{bmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ \vdots & \\ x'_n & y'_n \end{bmatrix}$$

Then find covariance matrix

$$C = \begin{bmatrix} x'_1 & x'_2 & \cdots & x'_n \\ y'_1 & y'_2 & \cdots & y'_n \end{bmatrix} \begin{bmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ \vdots & \\ x'_n & y'_n \end{bmatrix} / (n-1)$$

How is this useful?

p. 248: The axis passing through the points which minimizes the sum of the squared lengths of the perpendicular projection of the data points onto that axis is the principal axis

<u>Eigenvectors</u> provide the principal axes.

(135°)
axis of maximal projection length

axis of minimal projection length
(45°)

<u>Eigenvalues</u> indicate amount of information

0 value (or low) means that dimension not necessary