$n \times k \quad k \times n - \quad k \times m \qquad m \times n \qquad k \times n$

n = no of feature / dimension

m = user / training set

P.C.A (Principal Component Analysis)

use for dimensional reduction.

our algorithm is of this transpose

X = feature matrix

$$X = \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{array}\begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_n \\ - & - & - & - \\ - & - & - & - \end{bmatrix}$$

$x_1 \in R^n$

$$u = \begin{bmatrix} u_1 & \cdots & u_k \\ \theta_1 & & \theta_1 \\ \vdots & & \vdots \\ \theta_n & & \theta_n \end{bmatrix}$$

$n \times k$

Now we reduce dimension from n to k

$$X_{\text{feature matrix}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ \theta_1 & \theta_1 & & \theta_1 \\ \theta_2 & \theta_2 & & \theta_2 \\ \vdots & \vdots & & \vdots \\ \theta_n & \theta_n & & \theta_n \end{bmatrix}$$

$n \times m$

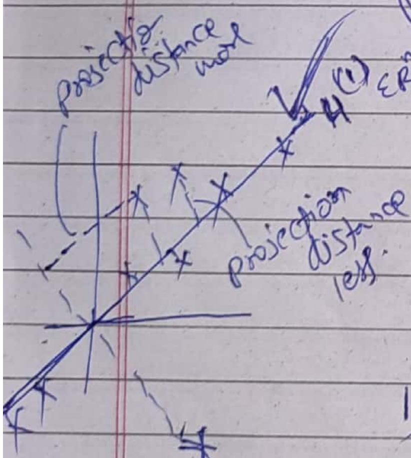
so, $x_1 \rightarrow z_1$

$$\begin{array}{c} z_1 \\ z_2 \\ \vdots \\ z_m \end{array}\begin{bmatrix} \theta_1 & - & - & \theta_k \\ & & & \\ \theta_1 & - & - & \theta_k \end{bmatrix}$$

$z_1 \in R^k$

$$\rightarrow \begin{bmatrix} z_1 & z_2 & \cdots & z_m \\ \theta_1 & \theta_1 & & \theta_1 \\ \theta_2 & \vdots & & \vdots \\ \theta_k & \theta_k & & \theta_k \end{bmatrix}$$

$k \times m$



so, PCA helps in finding the vector $u^{(1)} \in R^n$ on which to project the data minimize the projection error.

1) find K

2) Now, find K vectors $u^{(1)}, u^{(2)}, \ldots u^{(K)}$ over which we project the data

Ex, let we have 3 dimension data (x, y, z)

Now we want to reduce to two dimension $(K=2)$ x, y (u^1, u^2) Now we will project every data $(x_1, \ldots x_m)$ over x & y & these they will denoted by $(z_1, \ldots z_m)$

u is n dimensional vector and no of such u is K

$$\boxed{Z = U^T X}$$

$$\underset{k \times m}{Z} \quad \underset{(k \times n)(n \times m)}{U^T X}$$

going to original data X (n dimensional) from Z ($\in R^k$) reduced dimension

$$X_{approx} = U_{reduce} * Z$$

$$n \times K \quad K \times m$$

X_{approx} $n \times m$

finding U

To find k

1) calculating covariance matrix

(a) Σ (covariance) $= \frac{1}{m} \sum_{i=1}^{n} x^i (x^i)^T$

covariance matrix $= \frac{1}{m} * X * X^T$

(b) eigen vector

$[U, S, V] = svd(\text{covariance}(\Sigma))$

↑ singular value decomposition

b(i) $U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}_{n \times n}$

Here we needed 'k', 'u' vector so select 1st k columns

$U = U(:, 1:k);$

↑ reduced to k

$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_k \\ | & | & & | \end{bmatrix}_{n \times K}$

In starting given feature matrix feature normalizing is better so

$$\mu_j = \frac{1}{m} \sum_{i=1}^{n} x_j^i \quad 1 \le j \le n$$

$$X = \begin{array}{c} \begin{matrix} x_1 & x_2 & \cdots & x_m \end{matrix} \\ \begin{bmatrix} \theta_1 & \theta_1 & \cdots & \theta_1 \\ \theta_2 & \theta_2 & \cdots & \theta_2 \\ \vdots & \vdots & & \vdots \\ \theta_n & \theta_n & \cdots & \theta_n \end{bmatrix} \end{array} \rightarrow \begin{bmatrix} \text{sum}/m & \mu_1 \\ \text{2nd row sum}/m & \mu_2 \\ & \mu_3 \\ & \vdots \\ & \mu_n \end{bmatrix}$$

$$X' = \begin{array}{c} \begin{matrix} x_1 & x_2 \end{matrix} \\ \begin{bmatrix} \theta_1 - \mu_1, & \theta_1 - \mu_1 \\ \theta_2 - \mu_2 & \end{bmatrix} \end{array}$$

$\{z^1, z^2 - - - z^m\}$ $z^i \in R^k$ $k < n$

principal component analysis :

Reduce data from n dimension to k dimension

Compute covariance matrix.

① $\underset{(\text{sigma})}{\Sigma} = \frac{1}{m} \sum_{i=1}^{n} (x^{(i)}) (x^{(i)})^T$

Covariance matrix is calculate by (suppose we have matrix a, calculate avg of each feature which is column here) subtract it from original matrix result called deviation matrix (d). then take

$$\begin{array}{cc} & P \quad C \quad M \\ \begin{array}{c} 5 \\ \text{student} \end{array} & \begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \\ - & - & - \\ - & - & - \end{bmatrix} \end{array}$$

$\frac{sum(P)}{5}$ $\frac{sum(C)}{5}$ ..

② Calculat Eigen vector of covariance matrix.

$[u, S, V] = svd\ (\text{sigma})$

Singular value decomposition

given matrix, eigenvalue, eigenvector

$(A - \lambda I) X = 0$ — (i)

Identity matrix

① $|A - \lambda I| = 0$

calculat λ from here & put in (i) to get eigenvec

$$Z = \begin{bmatrix} | & & | \\ u^{(1)} & - - - & u^k \\ | & & | \end{bmatrix} \cdot x = \begin{bmatrix} - - (u^1)^T \\ (u^2)^T \\ \vdots \\ (u^k)^T \end{bmatrix} x$$

Scanned with CamScanner

LOKENATH
Page No.
Date

$$z_j^i = (u^j)^T x$$

Choosing K (no. of Principal Component)

Avg Squared Projection error :

$$\frac{1}{m}\sum_{i=1}^{m} \| x^i - x^i_{approx} \|^2$$

total variation in the data : $\frac{1}{m}\sum_{i=1}^{m} \| x^i \|^2$

choose K such that :

assign K=1 then 2, 3, 4 ... such that condition hold

$$\frac{\frac{1}{m}\sum_{i=1}^{m} \| x^i - x^i_{approx} \|^2}{\frac{1}{m}\sum_{i=1}^{m} \| x^i \|^2} \leq 0.01 \quad (1\%)$$

$[U, S, V] = svd(\text{sigma})$ ← covariance matrix

$d^T d$

U is matrix eigen vector

$$\begin{bmatrix} S_{11} & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 \\ 0 & 0 & 0 & S_{nn} \end{bmatrix}$$

- Non diagonal elements are 0.

it will also give K

for given K (select from begining across diagonal till K)

$$\left[1 - \frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{n} S_{ii}} \right] \leq 0.01$$

sum of all diagonal

better way to calc K:

$$\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{n} S_{ii}} \geq 0.99 \quad (99\% \text{ variance retained})$$

↑ K till we get this