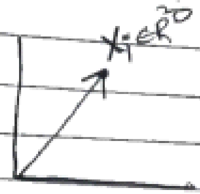
PCA formulation

Monday, September 28, 2020 11:45 PM

Date:

P. No:

Q. How to find the direction that preserves maximum variance?



For projecting $\vec{x_i}$ in any other direction, we have to multiply $\vec{x_i}$ with a unit vector in that direction

so basically when we say we want the direction which preserves maximum variation we want the unit vector in that direction $(\hat{u})$

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

⇒ 2 norm of unit vector is one $\|\hat{u}\| = 1$

Let's assume that our data is mean-centered. (Mean of each column is zero). Subtract mean from each column and make it mean-centered.

Variance of new data along $\hat{u}$ $= \frac{1}{N} \sum_{i=1}^{N} (x_i^T \hat{u})^2$

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

$$= u^T \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right)}_{S = \text{sample covariance Matrix of } X} \hat{u}$$

covariance $\frac{1}{N} \sum (x_i - \mu)(x_i - \mu)^T$

$\mu = 0$ in our case

So

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

$$\boxed{\sigma = \hat{u}^T S \hat{u}}$$

We want a $\hat{u}$ such than the term $\hat{u}^T S \hat{u}$ is maximized and at the same time $\hat{u}u = 1$ i.e. it is a unit vector

Each Principal component is a weigh...

Date:

P. No:

Objective function of PCA.

find $\hat{u} \longrightarrow \max \hat{u}^T S \hat{u}$ st. $\hat{u}^T u = 1$.

Constrained optimization problem. Lets introduce a Lagrangian

$$\max_{\hat{u}, \lambda} \hat{u}^T S \hat{u} - \lambda(\hat{u}^T u - 1)$$

Derivative wrt u and λ & equate to 0

$$\frac{\partial u^T S u}{\partial u} = 2 S \hat{u}$$

$$\therefore 2S\hat{u} - 2\lambda\hat{u} = 0.$$

$$\boxed{\begin{aligned} S\hat{u} &= \lambda\hat{u} \\ \frac{\partial}{\partial\lambda}\quad \hat{u}^T u &= 1 \end{aligned}}$$

This is where we go back to Eigen vector Analysis (see $S\hat{u} = \lambda\hat{u}$ form)

Property of eigen vector = They are unit length

find eigen vector of **S**. That will be $\hat{u}$

Once we find S then

$$\hat{u}_{max} = \hat{u}^T S \hat{u}$$

$$= \hat{u}^T \lambda u.$$

$$= \lambda \hat{u}^T u$$

$$\boxed{\hat{u}_{max} = \lambda}$$

$S\hat{u}$ was the variance along $\hat{u}$ & we see that it is λ

We will get D eigen vectors which will satisfy the equation as S is a D×D matrix

Projection in Direction $\begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_D \\ \lambda_1 & \lambda_2 & & \lambda_D \end{bmatrix}$.

The eigen vector corresponding to the largest Eigen value gives the direction of Maximal variance.

PCA is a linear algorithm.

$x_i \to z_i$ (Dimensionality Reduction)

$$\underset{}{z_i} = \underset{L\times D}{W^T}\ \underset{D\times 1}{x_i}$$

$z_i \to \hat{x}_i$ (Reconstruction)

$$\hat{x}_i = \underset{D\times L}{W'}\ \underset{L\times 1}{z_i}$$

when will Reconstruction be useful? → Compression

How to see loss of information due to PCA? Recover the data back & check the difference (shown in lab)

→ PCA for face Recognition.

Complexity of PCA (mainly Eigen value decomposition step) = $O(D^3)$ → D = no. of original features.