PCA

Tuesday, April 26, 2022 8:23 AM

Principal components

Given a set of data in $\mathbb{R}^d$ (Eucliden space of d-dimension), PCA provides sequence of best linear approximations of rank $q \le d$

let $x_1, x_2, \ldots, x_n$ be our observational data $X_{d \times n}$

Assuming a linear model $f(\lambda) = \mu + V_q \lambda$

μ – mean vector of size $d \times 1$

V_q – rank q matrix of size $d \times q$

λ – vector of size $q \times 1$

The vector λ is the parametric representation of the linear model f (a hyperplane)

define $\tilde{X} = X - \mu$ i.e. each observation (row) after mean subtraction

One approach to solve for V_q is orthogonalizing the covariance

Note that given two sets with zero mean

$$A = \{a_1, a_2, \ldots, a_n\} \quad B = \{b_1, b_2, \ldots, b_n\}$$

$$\bar{a} \qquad \bar{b}$$

The variance between A, B is $\sigma^2_{AB} = \frac{1}{n} \sum a_i b_i$

$$= \frac{1}{n} \bar{a} \bar{b}^T$$

Now for the matrix $\tilde{X}$ (note that it has mean zero) we can define the covariance as $C_X = \frac{1}{n} \tilde{X} \tilde{X}^T$ (size $n \times n$)

where the diagonal terms are variance, off diagonal terms are covariance

- large diagonal terms corresponds to interesting structure
- large off diagonal terms $\rightarrow$ redundancy

From the above, the solution to obtain a variance maximizing hyperplane is to diagonalize the variance after projection

define $\tilde{X} = V_q Y \quad (d \times n = (d \times q) \times (q \times n))$

where Y is a matrix of parametric representation in the linear model above (i.e. each row is a λ corresponding to $x \in \mathbb{R}^d$)

covariance of Y : $\quad C_Y = \frac{1}{n} Y Y^T$

$$= \frac{1}{n} (V_q^{-1} \tilde{X})(V_q^{-1} X)^T$$

$$= \frac{1}{n} (V_q^{-1} \tilde{X} \tilde{X}^T V_q^{-T})$$

$$= V_q^{-1} C_X V_q^{-T}$$

Note that C_X is a symmetric matrix thus decomposes as

$C_X = E^T D E$ where E is a eigen vector matrix

D is eigen value matrix

$$C_Y = V_q^{-1} E^T D E V_q^{-T}$$

To make C_Y diagonal (i.e. no covariance in Y), we can select

$$V_q = E^T$$

$$\Rightarrow C_Y = (E^{-T} E^T) D (E E^{-1}) = D$$

Now, we have the full model : $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \;, \; V_q = E^T$

Notes:

1. you can reconstruct any given sample x_i using the linear model $f(\lambda)$

by setting λ appropriately in $f(\lambda)$

2. Each column in V_q is called a principal component and it is of the same size as the original data point

3. if we arbitrarily set the λ to be unit vectors along a principal component, we can observe the changes to sample when traveled along that principal component.

4. The diagonal Eigen value matrix D also sorts principal components based on importance. Entries of D are refered to as explained Variance since they come from the covariance matrix C_X