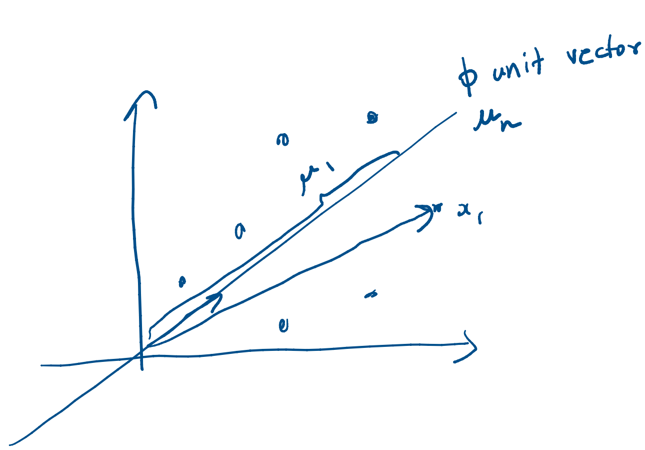
PCA

Tuesday, November 15, 2022 4:05 PM

lily

Steps

- Subtract the mean from *X*
- Calculate *Cov(X, X)*
- Calculate eigenvectors and eigenvalues of the covariance matrix
- Sort the eigenvectors according to their eigenvalues in decreasing ord
- Choose first k eigenvectors and that will be the new k dimensions
- Transform the original n-dimensional data points into k dimensions (= Projections with dot product)



$\phi^T\phi = 1$

$x_n^T \phi = \phi^T x_n$

$\hat{x}_n = \mu_n \phi$

$\hat{x}_n = \phi^T x_n \phi$

$\therefore \mu_n = \phi^T x_n$

$\mu_m = \phi_m^T x_n$

$\hat{x}_n = \mu_m \phi_m$

$$\mathcal{L}(\phi) = \sum_n \| x_n - \mu_n \phi \|^2$$

$$= \sum_n (x_n - \mu_n \phi)^T (x_n - \mu_n \phi)$$

$$= \sum_n x_n^T x_n - x_n^T \mu_n \phi - \mu_n \phi^T x_n + \mu_n^2 \overbrace{\phi^T \phi}^{1}$$

$$= \sum_n x_n^T x_n - 2\mu_n \phi^T x_n + \mu_n^2$$

$$= \sum_n x_n^T x_n - \mu_n^2$$

$$\mathcal{L}(\phi) = \underbrace{\sum_n x_n^T x_n}_{\text{constant}} - N\left(\frac{1}{N}\sum_n \mu_n^2\right)$$

Step 1.

Calculate error

Let's min error

Ops.. min error means max projected variance

Minimizing error means maximizing variance on projected axis

... is also 0.

Minimizing error means maximizing variance on projected ...

If mean of x_n is 0, mean of u_n is also 0.

$$u^{(n)} = \begin{bmatrix} \phi_1^T x^{(n)} \\ \phi_2^T x^{(n)} \\ \vdots \\ \phi_M^T x^{(n)} \end{bmatrix} = \begin{bmatrix} - \phi_1^T - \\ - \phi_2^T - \\ \vdots \\ - \phi_M^T - \end{bmatrix} \begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_D^{(n)} \end{bmatrix}$$

How variant it is on projected axis / surface

$(M \times 1)$ $\quad (M \times D) \quad (D \times 1)$

$$= \phi^T x^{(n)}$$

$$\phi = \begin{bmatrix} | & | & & | \\ \phi_1 & \phi_2 & \cdots & \phi_M \\ | & | & & | \end{bmatrix}, \quad x^{(n)} = \begin{bmatrix} x_1^{(n)} \\ \vdots \\ x_D^{(n)} \end{bmatrix}$$

$(D \times M) \quad (D \times 1)$

$$\sigma_{u_1}^2 = \frac{1}{N}\sum_{n=1}^{N} (u_1^{(n)} - \bar{u}_1)^2 \rightarrow 0$$

$$= \frac{1}{N}\sum_{n=1}^{N} (\phi_1^T x^{(n)})^2$$

$$= \frac{1}{N}\sum_{n=1}^{N} (\phi_1^T x^{(n)}) (\phi_1^T x^{(n)})^T$$

$$= \frac{1}{N}\sum_{n=1}^{N} (\phi_1^T x^{(n)}) ((x^{(n)})^T \phi_1)$$

$$= \phi_1^T \left[\frac{1}{N}\sum_{n=1}^{N} (x^{(n)}) (x^{(n)})^T \right] \phi_1$$

Step 2.

max projected variance

oops. it means finding vector directions that make x variance max

2dimensional

$$X \begin{bmatrix} x_1 \cdots n \\ x_2 \cdots n \end{bmatrix}$$

Cov

$$\begin{bmatrix} x_1^{(n)} - \bar{x}_1 & x_1^{(n)} - \bar{x}_2 \\ x_2^{(n)} - \bar{x}_1 & x_2^{(n)} - \bar{x}_2 \end{bmatrix}$$

If mean not zero

$$\sigma_{u_1}^2 = \phi_1^T \left[\overbrace{\frac{1}{N}\sum_{n=1}^{N} (x^{(n)} - \bar{x}_{Cov})(x^{(n)} - \bar{x}_{Cov})^T}^{\text{Cov} \leftarrow \text{matrix}} \right] \phi_1$$

D×D for all

$$\max \sigma_u \Rightarrow \min - \sigma_u$$

Step 3.

nk then diff

$\max \sigma_u \Rightarrow \min -\sigma_u$

$$J(\phi_1) = -\sigma_u^2 + r(\|\phi_1\|^2 - 1)$$

$$= -\phi_1^T Cov \phi_1 + r(\|\phi_1\|^2 - 1)$$

$$\frac{\partial J(\phi_1)}{\partial \phi_1} = -\cancel{2} Cov \phi_1 + \cancel{2} r \phi_1 = 0$$

$$Cov \phi_1 = r \phi_1$$

Step 3.

ok then diff

projected variance with vector directions

oops... constraints careful

$$\phi_1^T Cov \phi_1 = \phi_1^T r \phi_1$$

$$\sigma_u^2 = r$$

Maximize variance

$\therefore r$ max

So eigenvalue max, eigen vector max

Step 4.

Cool. Now, which eigenvectors / eigen values pairs should we use?

Calculate cumulative contribution rate $\lambda = r$

$$\frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{D} \lambda_i} > 0.8 \sim 0.9 \quad \text{ideally}$$

Original dimension D

d - no. of dimension to reduce to

Step 5.

OK, but only one value? How much can / should we really reduce?
