Machine Learning

Principal Components Analysis

1. Basics

variance:

$$D\xi := E(\xi - E\xi)^2 = E\xi^2 - (E\xi)^2$$

normalization:

$$\frac{\xi - E\xi}{\sqrt{D\xi}}$$

covariance:

$$egin{aligned} \cot(\xi,\eta) := E[(\xi-E\xi)(\eta-E\eta)] \ = E\xi\eta-E\xi\cdot E\eta \end{aligned}$$

correlation:

$$ho(\xi,\eta) := rac{\mathrm{cov}(\xi,\eta)}{\sqrt{D\xi \cdot D\eta}}$$

SVD:

$$M = U \Sigma V^{ op} ext{ where } U^{ op} U = I, V^{ op} V = I$$
 $M^{ op} M V = V \Sigma^2$ $M M^{ op} U = U \Sigma^2$

2. Model

input:

$$\{x^{(i)}; i=1,\ldots,m\}, x^{(i)} \in \mathbb{R}^n$$

normalization:

$$x^{(i)} = x^{(i)} - rac{1}{m} \sum_{i=1}^m x^{(i)} \ x^{(i)}_j = rac{x^{(i)}_j}{\sqrt{rac{1}{m} \sum_{i=1}^m (x^{(i)}_j)^2}}$$

objective function:

$$egin{aligned} J(u) &= rac{1}{m} \sum_{i=1}^m ((x^{(i)})^ op u)^2 = rac{1}{m} \sum_{i=1}^m u^ op x^{(i)} (x^{(i)})^ op u \ &= u^ op \left(rac{1}{m} \sum_{i=1}^m x^{(i)} (x^{(i)})^ op
ight) u = u^ op \Sigma u \end{aligned}$$

optimization:

$$\max J(u)$$
 s.t $u^{\top}u = 1$

solution:

u is the principal eigenvector of $\Sigma = rac{1}{m} \sum_{i=1}^m x^{(i)} (x^{(i)})^ op = rac{1}{m} X^ op X$.

The k-th component:

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X u_{(s)} u_{(s)}^{ op} \ u_{(k)} = \operatorname{argmax}_{||u||=1} rac{1}{m} ||\hat{X}_k u||^2$$

the eigenvectors of Σ :

$$X^ op = egin{bmatrix} | & | & | & | \ x^{(1)} & x^{(2)} & \cdots & x^{(n)} \ | & | & | \end{bmatrix}$$

$$\mathrm{SVD}: X^\top = USV^\top$$

$$U=egin{bmatrix}dash u_{(1)} & u_{(2)} & \cdots & u_{(n)} \dash dash ash ash ash ash dash ank dash dash dash dash ank ank ank ank ank ank d$$

output:

$$y^{(i)} = egin{bmatrix} u_{(1)}^ op x^{(i)} \ u_{(2)}^ op x^{(i)} \ dots \ u_{(k)}^ op x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

recover:

$$\hat{x} = U egin{bmatrix} ilde{y}_1 \ ilde{y}_k \ 0 \ ilde{dots} \ 0 \end{bmatrix} = \sum_{i=1}^k u_{(i)} ilde{y}_i.$$

3. TODO

- 1. Optimization
- 2. Calculating the SVD