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:

$$\begin{aligned} \mathrm{cov}(\xi, \eta) & := E[(\xi - E\xi)(\eta - E\eta)] \\ & = E\xi\eta - E\xi \cdot E\eta \end{aligned}$$

correlation:

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

SVD:

$$\begin{gathered} M = U\Sigma V^\top \text{ where } U^\top U = I, V^\top V = I \\ M^\top M V = V\Sigma^2 \\ M M^\top U = U\Sigma^2 \end{gathered}$$

2. Model

input:

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

normalization:

$$x^{(i)} = x^{(i)} - \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$$

$$x_j^{(i)} = \frac{x_j^{(i)}}{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)})^2}}$$

objective function:

$$J(u) = \frac{1}{m}\sum_{i=1}^{m}((x^{(i)})^\top u)^2 = \frac{1}{m}\sum_{i=1}^{m} u^\top x^{(i)}(x^{(i)})^\top u$$
$$= u^\top \left(\frac{1}{m}\sum_{i=1}^{m} x^{(i)}(x^{(i)})^\top\right) u = u^\top \Sigma u$$

optimization:

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

solution:

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

The k-th component:

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X u_{(s)} u_{(s)}^\top$$

$$u_{(k)} = \text{argmax}_{||u||=1} \frac{1}{m}||\hat{X}_k u||^2$$

the eigenvectors of Σ:

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

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

$$U = \begin{bmatrix} | & | & & | \\ u_{(1)} & u_{(2)} & \cdots & u_{(n)} \\ | & | & & | \end{bmatrix}$$

output:

$$y^{(i)} = \begin{bmatrix} u_{(1)}^\top x^{(i)} \\ u_{(2)}^\top x^{(i)} \\ \vdots \\ u_{(k)}^\top x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

recover:

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

3. TODO

1. Optimization
2. Calculating the SVD