Notes on Principal Component Analysis (PCA)

Arshak Minasyan

January 2016

1 Basics

1.1 Singular value decomposition (SVD)

SVD is defined for arbitrary matrix $X \in \mathbb{M}^{n\times d}(\mathbb{R})$ ($\mathbb{R}$ could be easily replaced by $\mathbb{C}$) and let $r = \operatorname{rank} X$, then

$$\underbrace{X}_{n\times d} = \underbrace{U}_{n\times r}\underbrace{\Sigma}_{r\times r}\underbrace{V^T}_{r\times d}, \tag{1}$$

where

$$U = \begin{pmatrix} \vdots & & \vdots \\ u_1 & \dots & u_r \\ \vdots & & \vdots \end{pmatrix} \quad \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r) \quad V = \begin{pmatrix} \vdots & & \vdots \\ v_1 & \dots & v_d \\ \vdots & & \vdots \end{pmatrix} \tag{2}$$

with $u_i \in \mathbb{R}^n$ for $i = 1, \dots, r$, $v_i \in \mathbb{R}^r$ for $i = 1, \dots, d$ and $\sigma_i > 0$ for $i = 1, \dots r$. Moreover, matrices U and V are unitary, i.e. $U^T U = I, V^T V = I$.

1.2 Eckart-Young theorem

2 PCA

Suppose we are given a cluster $\mathcal{L}_n = \{x_1, \dots, x_n\}$, where $x_i \in \mathbb{R}^d$ for $i = 1, \dots, n$. We define matrix X as follows

$$\underbrace{X}_{n\times d} = \begin{pmatrix} x_{11} & \dots & \dots & x_{1d} \\ \vdots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ x_{n1} & \dots & \dots & x_{nd} \end{pmatrix} \tag{3}$$

Then, we center our data by subtracting the sample mean from all observations, i.e. $x_i := x_i - \overline{x}$, where $\overline{x} = \dfrac{1}{n}\sum_{i=1}^n x_i \in \mathbb{R}^d$. So the sample covariance matrix could be written in this way

$$S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} X^T X \tag{4}$$

Thus, using singular value decomposition for matrix $X = U\Sigma V^T$ one can obtain

$$X^T X = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T \implies \underbrace{\frac{1}{n} X^T X}_{S} v_i = \frac{\sigma_i^2}{n} v_i, \tag{5}$$

which means that the pairs $\left(\frac{\sigma_i^2}{n}, v_i\right)$ are the eigenvalue-eigenvector pairs of matrix S.

The i-th principal component of matrix S is

$$\underbrace{z_i}_{n\times 1} = \underbrace{X}_{n\times d}\underbrace{v_i}_{d\times 1} \implies z_i = U\Sigma V^T v_i \implies z_i = \sigma_i u_i \tag{6}$$

or, aggregating all the principal components one can get

$$\underbrace{Z}_{n\times r} = \underbrace{X}_{n\times d}\underbrace{V}_{d\times r} = \underbrace{U}_{n\times r}\underbrace{\Sigma}_{r\times r}, \tag{7}$$

where $Z = \begin{pmatrix} \vdots & & \vdots \\ z_1 & \dots & z_r \\ \vdots & & \vdots \end{pmatrix}$.

2.1 Kernel PCA

Take a look at HSE french guy, at CS department.

2.2 Probabilistic PCA

Vetrov SHAD.

3 Robust PCA