MAT 653: Statistical Simulation

Instructor: Dr. Wei Li   Scribe: Waleed A. Raja

Sep 16th, 2021

Principal Component

X: $n \times p$ matrix, X is a data matrix, n is the number of cases, and p is the number of features.

Example: n=3 and p=2

$$X = \begin{bmatrix} * & * \\ * & * \\ * & * \end{bmatrix}$$

$$X = U\Sigma V^T$$

X is a $n \times p$ matrix, Σ is a $n \times p$ matrix, and V^T is a $p \times p$ matrix.

Here $V_j = j^{th}$ column of V is called j^{th} principal component direction of X.

$$\|XV_j\| = \sigma_j,$$

where σ_j are singular values of X.

Each element in V_j are called principal component loadings.

$$XV_j = Z_j$$

where Z_j is called the j^{th} principal component of X. Elements in Z_j are called Principal Component scores.

$$\begin{aligned} Z_j &= (Z_{1j}, Z_{2j}, \dots, Z_{nj}) \\ \text{where } Z_{ij} &= X_{[i,.]}V_j \\ &= \frac{X_{[i,.]}V_j}{\langle V_j, V_j \rangle} \end{aligned}$$

$\frac{X_{[i,.]}V_j}{\langle V_j, V_j \rangle}$ is the coefficient of projecting $X_{[i,.]}$ onto span of $\{V_j\}$.

Sample variance of $Z_j = XV_j$, $\frac{\|XV_j\|^2}{n-1} = \frac{\sigma_j^2}{n-1}$

Since $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2$,$Z_1 = XV_1$ has the largest sample variance among all normalized linear combinations of columns of X; Z_j, $j \geq 2$, has maximum sample variance subject to being orthogonal to the earlier ones. $Z_2 \perp Z_1, Z_3 \perp (Z_2, Z_1), \dots$.

Check:

$$\begin{aligned} Cov(Z_1, Z_2) &= \frac{Z_1^T Z_2}{n-1} \\ &= \frac{(XV_1)^T XV_2}{n-1} \\ &= \frac{V_1^T X^T XV_2}{n-1} \\ &= \frac{(V_1^T V)\sum\sum(V^T V_2)}{n-1} \\ &= 0 \end{aligned}$$

Since $XV_j = Z_j$, $X[V_1, \ldots, V_p] = [Z_1, Z_2, \ldots, Z_p]$.

One application for the principal components is the principal component regression:

Least Square Problem

$\min_{\beta} \|X\beta - y\|^2$

Principal Components Regression

$\min_{\theta} \|Z\theta - y\|^2$

By construction, the first principal component will contain the most information about the data, the subsequent principal components contain less and less information about the data. Therefore, the first few principal components $Z_1, Z_2, \ldots, Z_k$ $(k < p)$ can be used as predictors in lieu of the original set of all predictors in X.

Multivariate Normal

If $Z_1, Z_2, \ldots, Z_p \overset{iid}{\sim} \mathcal{N}(0, 1)$,

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

Then, $Z \sim \mathcal{N}(0, I)$

Also, by the well known property of multivariate normal distribution, $\mu + AZ \sim \mathcal{N}(\mu, AA^T)$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1p}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2p}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}^2 & \sigma_{p2}^2 & \cdots & \sigma_p^2 \end{bmatrix}$$

AA^T is the variance/covariance matrix. It is a symmetric matrix.

Goal: To generate W (multivariate), such that, $W \sim \mathcal{N}(\mu, \sum)$

Example: $p = 2$

$$W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}$$

$$\sum = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}$$

Take Cholesky decomposition of $\sum = U^T U$. In R, $\mu + U^T Z \sim \mathcal{N}(\mu, U^T U)$