Class 9 Review Notes

AI & Machine Learning

Fall 2023 - Laurie Ye

9 Principle Components Anaysis

9.1 Notation and Preliminaries

Y is a random vector:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}, \quad E(Y) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu$$

The variance covariance matrix for Y is

$$\begin{aligned} \text{var}(Y) &= \begin{bmatrix} \text{var}(Y_1) & \text{cov}(Y_1, Y_2) & \cdots & \text{cov}(Y_1, Y_p) \\ \text{cov}(Y_2, Y_1) & \text{var}(Y_2) & \cdots & \text{cov}(Y_2, Y_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(Y_p, Y_1) & \text{cov}(Y_p, Y_2) & \cdots & \text{var}(Y_p) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} \sigma_{ij} \end{bmatrix} = \Sigma \end{aligned}$$

The variance-covariance matrix is symmetric.

$$\Sigma = \Sigma' = \Sigma^t, \quad \sigma_{ij} = \sigma_{ji}, \quad \sigma_{ii} = \sigma_i^2$$

9.2 Fact 1

For a conforming matrix A (i.e., A is $l \times p$) and vector b (i.e., b is a length p) of scalars (i.e., non-random):

$$E(AY + b) = AE(Y) + b = A\mu + b$$

$$\text{var}(AY + b) = A\text{var}(Y)A' = A\Sigma A'$$

For scalar vectors a and b (of length p):

$$E(a'Y + b) = a'E(Y) + b = a'\mu + b$$

$$\text{var}(a'Y + b) = a'\text{var}(Y)a = a'\Sigma a$$

9.3 Fact 2 (Eigen Decomposition)

The variance-covariance matrix can be decomposed as follows:

$$\Sigma = P\Lambda P' \quad \text{— implies —} \quad P'\Sigma P = \Lambda$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_p \end{bmatrix} \quad \text{(Diagonal)}$$

and P is orthonormal, meaning that

$$P'P = I = PP'$$

where I is the identity matrix.

9.4 Eigenvalues

Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of Σ. It is customary to order the columns and rows of $P\Lambda P'$ so that the eigenvalues are in order from largest to smallest:

$$\max(\lambda_1, \ldots, \lambda_p) = \lambda_1$$

$$\min(\lambda_1, \ldots, \lambda_p) = \lambda_p$$

9.5 Eigenvectors

The columns of P are the eigenvectors of Σ: $P = [P_1 \; P_2 \; \ldots \; P_p]$ where $P_j = \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{pj} \end{bmatrix}$.

Note that $P_j'P_j = 1$ and $P_j'P_i = 0$ for $i \neq j$.

P_j corresponds to λ_j in that $P_j'\Sigma P_j = \lambda_j$.

9.6 Fact 3

Suppose $\mathbf{a}$ is of length 1 (i.e., $\mathbf{a}'\mathbf{a} = 1$). Then

$$\max_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \text{var}(\mathbf{a}'\mathbf{Y}) = \max_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \mathbf{a}'\Sigma\mathbf{a} = \lambda_1$$

The solution $\mathbf{a}$ is $\mathbf{P}_1$:

$$\text{arg max}_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \text{var}(\mathbf{a}'\mathbf{Y}) = \text{arg max}_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \mathbf{a}'\Sigma\mathbf{a} = \mathbf{P}_1$$

Comments:

- Very good numerical routines exist for computing the eigenvector-eigenvalue decomposition. In R, the function is called `eigen()` and is a part of base R (no special package necessary).

Fact 3 (cont.)

Suppose $\mathbf{a}$ is of length 1 (i.e., $\mathbf{a}'\mathbf{a} = 1$) and is also orthogonal to eigenvectors 1 to $j-1$; that is

$$\mathbf{a}'\mathbf{a} = 1 \quad \text{and} \quad \mathbf{a}'\mathbf{P}_1 = 0, \quad \mathbf{a}'\mathbf{P}_2 = 0, \ldots, \quad \mathbf{a}'\mathbf{P}_{j-1} = 0$$

then

$$\max_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \text{var}(\mathbf{a}'\mathbf{Y}) = \max_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \mathbf{a}'\Sigma\mathbf{a} = \lambda_j$$

The solution $\mathbf{a}$ is now $\mathbf{P}_j$:

$$\text{arg max}_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \text{var}(\mathbf{a}'\mathbf{Y}) = \text{arg max}_{\mathbf{a} \text{ s.t. } \mathbf{a}'\mathbf{a}=1} \mathbf{a}'\Sigma\mathbf{a} = \mathbf{P}_j$$

9.7 The Data Matrix Y

All of the discussions so far has related to the random vector $\mathbf{Y}$. When I observe a sample of observations from this random vector, I generally store them as rows in a data matrix:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1' \\ \mathbf{Y}_2' \\ \vdots \\ \mathbf{Y}_n' \end{bmatrix}$$

where $\mathbf{Y}_i$ is the ith observation from the random vector $\mathbf{Y}$.

9.8 Principal Components

The first principal component is the linear combination of the random vector $\mathbf{Y}$ (i.e., a linear combination of the variables) that "explains" the maximum amount of variance of the data:

$$\mathbf{Y}\mathbf{P}_1 = \begin{bmatrix} \mathbf{Y}_1'\mathbf{P}_1 \\ \mathbf{Y}_2'\mathbf{P}_1 \\ \vdots \\ \mathbf{Y}_n'\mathbf{P}_1 \end{bmatrix} = \text{the 1st principal component}$$

The other principal components are similarly defined.

We can compute all of the principal components at once as a matrix:

$$\mathbf{Y}_{n\times p}\mathbf{P}_{p\times p} = [\mathbf{PC}]_{n\times p}$$

The ith column of $\mathbf{PC}$ is the ith principal component. This implies (you can do backwards)

$$\mathbf{Y}_{n\times p} = [\mathbf{PC}]_{n\times p}\mathbf{P'}_{p\times p}$$

9.9 Dimension Reduction

Principal components can be used to reduce the dimension by just not using the full set. Instead use some number of the most important eigenvectors. Suppose I use the first k eigenvectors, then

$$\mathbf{Y}_{n\times p}\mathbf{P}^*_{p\times k} = [\mathbf{PC}]^*_{n\times k}$$

9.10 Data "Reconstruction"

Once the dimension has been reduced by using a reduced number of the most important principal components, the data set can be reconstructed. If $\mathbf{P}^*$ is based on the first k eigenvectors, then the data can be "reconstructed" using:

$$\mathbf{Y}^*_{n\times p} = [\mathbf{PC}]^*_{n\times k}[\mathbf{P}^*]'_{k\times p}$$

9.11 Reference

Goizueta Business School-Emory University: Professor George S. Easton