ABE 6933 SML, Fall 2020
Supplementary Materials for SVD/PCA

Instructor: Dr. Nikolay Bliznyuk

11 November, 2020

Singular Value Decomposition (SVD)

Let X be any $n \times p$ matrix where $n \geq p$.

The SVD of X is the representation $X = UDV^T = \sum_{i=1}^{p} d_j U_j V_j^T$

$$X = [U_1, \ldots, U_p] \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_p \end{bmatrix} \begin{bmatrix} V_1^T \\ \vdots \\ V_p^T \end{bmatrix}, \text{ where}$$

- U is a $n \times p$ matrix with columns $U_1, \ldots, U_p$ such that $U^T U = I_p$.
- V is a $p \times p$ matrix with columns $V_1, \ldots, V_p$ such that $V^T V = I_p$.
- D is a $p \times p$ diagonal matrix such that $D_{ii} = d_i$ and $d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$ are *singular values*. $D_{ij} = 0$ if $i \neq j$.
- If there is $r < p$ such $d_r > 0$ and $d_{r+1} = 0$, then $r = \text{rank}(X)$.
- $\widetilde{X} = \sum_{i=1}^{m} d_j U_j V_j^T$ is the best rank-m approximation of X that minimizes $\|X - \widetilde{X}\|_F^2 = \sum_{i,j}(X_{ij} - \widetilde{X}_{ij})^2$ over all rank-m matrices.

Relationship between the SVD and PCA, I

- SVD is the algorithm for obtaining PC vectors and scores.
- Here, X is the design matrix without the column of ones.
- As a pre-processing step for PCA, it is customary to center the columns of X, as well as (in most cases) scale the columns of X so that they have unit variance (this will correspond to the "correlation" version of the PCA, rather than the "covariance" version). Then the SVD is applied to the pre-processed X.
- PC vectors of loadings are given by the columns of V, i.e., V_j is the jth PC loading vector (ϕ_j in the ISLR notation)
- The matrix of PC scores Z is obtained as $Z = XV = UDV^TV = UDI_p = UD$; i.e., $(UD)_{ij}$ is the jth PC score for the ith observation, and $Z_j = d_jU_j$ is the vector of the jth PC scores for the entire dataset.

Relationship between the SVD and PCA, II

- $Z_j = d_j U_j$ is the vector of the jth PC scores for the entire dataset.
- Columns of Z are orthogonal, i.e., $Z_j^T Z_k = d_j d_k U_j^T U_k = 0$ if $j \neq k$ and $Z_j^T Z_j = d_j^2$.
- Since the columns of X have been centered, the columns of Z are also centered (automatically). Hence
- $Var(Z_j) = \frac{1}{n} \sum_{i=1}^{n} (Z_{ij} - \bar{Z}_j)^2 = \frac{1}{n} \sum_{i=1}^{n} Z_{ij}^2 = \frac{1}{n} Z_j^T Z_j = d_j^2 / n$.
- $Cov(Z_j, Z_k) = \frac{1}{n} \sum_{i=1}^{n} (Z_{ij} - \bar{Z}_j)(Z_{ik} - \bar{Z}_k) = \frac{1}{n} \sum_{i=1}^{n} Z_{ij} Z_{ik} = \frac{1}{n} Z_j^T Z_k = 0$ if $j \neq k$.

Relationship between the SVD and PCA, III

- $S = \frac{1}{n} X^T X$ is the (sample) covariance matrix for the rows of X.
- The overall variability associated with predictors in X is $\sum_{j=1}^{p} Var(X_j) = tr(S)$.
- For a square matrix A, the trace of A is $tr(A) = \sum_{j=1}^{p} A_{jj}$.
- Notice that if A is $n \times p$ and B is $p \times n$, then $tr(AB) = tr(BA)$.
- Hence $tr(S) = tr(X^T X/n) = \frac{1}{n} tr(VD^T U^T UDV^T) = \frac{1}{n} tr(VDDV^T) = \frac{1}{n} tr(DDV^T V) = \frac{1}{n} tr(DDI_p) = \sum_{j=1}^{p} d_j^2/n = \sum_{j=1}^{p} Var(Z_j)$.

PCA/SVD and the Spectral Decomposition

- Recall $S = \frac{1}{n} X^T X$ is the covariance matrix for the rows of X.
- Plug in $X = UDV^T$ and simplify
  $S = \frac{1}{n} V D^T U^T U D V^T = V \Lambda V^T$ is the spectral decomposition of S, where $\Lambda = (\frac{1}{n} DD)$ is diagonal.
- Suppose W is a random vector with the covariance matrix S, e.g., $W \sim Mult.Normal(0, S)$. Then $Var(a^T W) = a^T S a$ is variation of W in the direction a whenever $\|a\|_2 = 1$.
- Specifically, $Var(V_j^T W) = V_j^T V^T \Lambda V^T V = \Lambda_{jj} = d_j^2/n$ is the variation of W along the direction of the j PC vector V_j.
- Notice $Cov(V_j^T W, V_k^T W) = V_j^T V \Lambda V^T V_k = 0$ if $j \neq k$.
- Hence PCA decomposes the overall variation in W additively into the variation along the orthogonal directions given by the PC vectors $V_1, \ldots, V_p$.