STA 141C - Big Data & High Performance Statistical Computing **Spring 2022**

Lecture 10: Introduction to singular value decomposition

Lecturer: Bo Y.-C. Ning *February 8, 2022*

Disclaimer: *My notes may contain errors, distribution outside this class is allowed only with the permission of the Instructor.*

Announcement

- HW2 due next Wednesday

Last time

- PageRank problem
- Iterative method

Today

- Singular value decomposition (SVD)
- Principal component analysis (PCA)

1 Review of eigenvalue and eigenvector

Assume $A \in \mathbb{R}^{n\times n}$ is a square matrix

- Eigenvalues are defined as roots of the characteristic equation $\det(\lambda I_n - A) = 0$
- If λ is an eigenvalue of A, then there exist non-zero $x, y \in \mathbb{R}^n$ such that $Ax = \lambda x$ and $y'A = \lambda y'$, x is the (column) eigenvector and y is the row eigenvector of A associated with the eigenvalue λ
- A is singular if and only if it has at least one 0 eigenvalue.
- Eigenvectors associated with distinct eigenvalues are linearly independent
- Eigenvalues of an upper or lower triangular matrix are its diagonal entries: $\lambda = a_{ii}$
- Eigenvalues of an idempotent matrix are either 0 or 1
- In most statistical applications, we deal with eigenvalues/eigenvectors of symmetric matrices. The eigenvalues and eigenvectors of a real symmetric matrix are real.
- Eigenvectors associated with distinct eigenvalues of a symmetry matrix are orthogonal.

- Eigen-decompostion of a symmetric matrix: $A = U\Lambda U'$, where
  - $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$
  - Columns of U are the eigenvectors which are mutually orthonormal
  - A real symmetric matrix is positive semidefinite (positive definite) if and only if all eigenvalues are nonnegative (positive).
  - $\text{tr}(A)$ (a square matrix not require to be symmetric), $\text{tr}(A) = \text{tr}(U\Lambda U') = \text{tr}(U'U\Lambda) = \text{tr}(\Lambda) = \sum_i \lambda_{ii}$

2 Review of singular value decomposition (SVD)

For a rectangular matrix $A \in \mathbb{R}^{m\times n}$, let $p = \min\{m, n\}$, then we have the SVD

$$A = U\Sigma V',$$

where $U = (u_1, \ldots, u_m)$ and $V = (v_1, \ldots, v_n)$ are orthogonal matrices and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$ is a diagonal matrix such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. σ_is are called the *singular values*, u_is are the left singular vectors and v_is are the right singular vectors.

The matrix Σ is not a square matrix, one can define thin SVD, which factorizes A as

$$A = U_n\Sigma_n V' = \sum_{i=1}^{n} \sigma_i u_i v_i',$$

where $U_n \in \mathbb{R}^{m\times n}$, $U_n'U_n = I_n$, $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$. This is for $m > n$, if $m < n$, then we let $V \in \mathbb{R}^{m\times n}$,

The following properties are useful: for $\sigma(A) = (\sigma_1, \ldots, \sigma_p)'$, the rank of A is the number of nonzero singular values denoted as $\|\sigma(A)\|_0$. The Frobenius norm of A, $\|A\|_F = (\sum_{i=1}^p \sigma_i^2)^{1/2} = \|\sigma(A)\|_2$, and the spectrum norm of A, $\|A\|_2 = \sigma_1 = \|\sigma(A)\|_\infty$. Using the fact that U, V are both orthogonal matrices

$$A'A = V\Sigma U'U\Sigma V' = V\Sigma^2 V',$$
$$AA' = U\Sigma V'V\Sigma U' = U\Sigma^2 U'$$

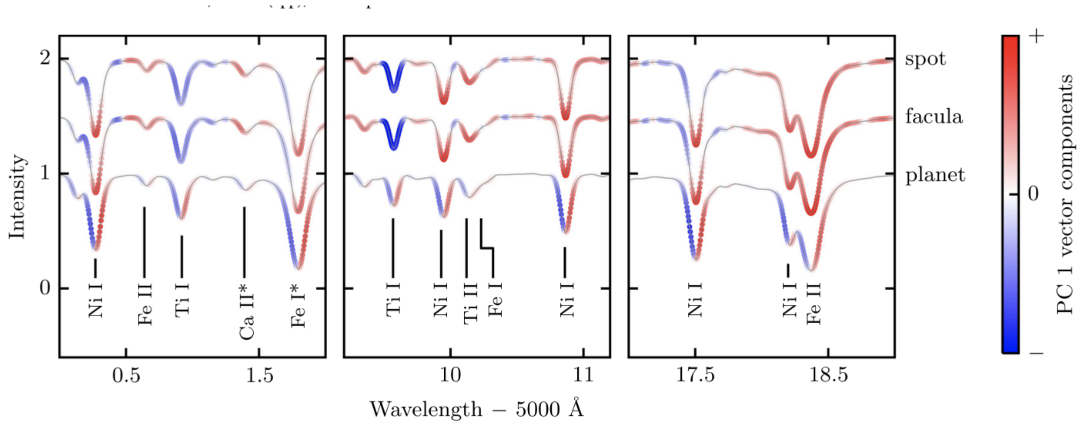
Last, the eigen-decomposition for a real symmetric matrix is $B = W\Lambda W'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, which is the SVD of B.

3 Applications for SVD

1. Principal component analysis and dimension deduction.

Let $X \in \mathbb{R}^{n\times p}$ be a centered data matrix, perform SVD on $X = U\Sigma V'$. The linear combinations $\tilde{x}_i = Xv_i$ are the principal components (PCs) with variance σ_i^2.

Dimension deduction: reduce dimensionality p to $q \leq p$ and use the first few PCs $\tilde{x}_i, \ldots, \tilde{x}_q$ in downstream analysis. Used in medical studies, astronomy, etc.



Davis et al. (2017). *Insights on the Spectral Signatures of Stellar Activity and Planets from PCA*. The Astrophysical Journal.