# Seminar Series on Linear Algebra for Machine Learning Part 5: Singular Value Decomposition and Principal Component Analysis


Dr. Ceni Babaoglu

Data Science Laboratory
Ryerson University

*cenibabaoglu.com*

# Overview

# Spectral Decomposition

- An $n \times n$ symmetric matrix $A$ can be expressed as the matrix product
$$A = PDP^T$$
where $D$ is a diagonal matrix and $P$ is an orthogonal matrix.
- The diagonal entries of $D$ are the eigenvalues of $A$, $\lambda_1, \lambda_2, \cdots, \lambda_n$, and the columns of $P$ are associated orthonormal eigenvectors $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n$.

# Spectral Decomposition

- The expression

$$A = PDP^T$$

is called the *spectral decomposition* of $A$. We can write it as

$$A = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix}$$

# Spectral Decomposition

- The matrix product $DP^T$ gives

$$DP^T = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{x}_1^T \\ \lambda_2 \vec{x}_2^T \\ \vdots \\ \lambda_n \vec{x}_n^T \end{bmatrix}$$

$$A = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 \vec{x}_1^T \\ \lambda_2 \vec{x}_2^T \\ \vdots \\ \lambda_n \vec{x}_n^T \end{bmatrix}$$

# Spectral Decomposition

- We can express $A$ as a linear combination of the matrices $\vec{x}_j\vec{x}_j^T$, and the coefficients are the eigenvalues of $A$,

$$A = \sum_{j=1}^{n} \lambda_j \vec{x}_j \vec{x}_j^T = \lambda_1 \vec{x}_1 \vec{x}_1^T + \lambda_2 \vec{x}_2 \vec{x}_2^T + \cdots + \lambda_n \vec{x}_n \vec{x}_n^T.$$

# Singular Value Decomposition

Singular Value Decomposition is based on a theorem from linear algebra which says the following:
a rectangular matrix $A$ can be broken down into the product of three matrices:

- an orthogonal matrix $U$,
- a diagonal matrix $S$,
- the transpose of an orthogonal matrix $V$.

# Singular Value Decomposition

Let $A$ be an $m \times n$ real matrix. Then there exist orthogonal matrices $U$ of size $m \times m$ and $V$ of size $n \times n$ such that

$$A = USV^T$$

where $S$ is an $m \times n$ matrix with nondiagonal entries all zero and $s_{11} \geq s_{12} \geq \cdots \geq s_{pp} \geq 0, \quad p = \min\{m, n\}.$

- the diagonal entries of $S$ are called the *singular values* of $A$,
- the columns of $U$ are called the *left singular vectors* of $A$,
- the columns of $V$ are called the *right singular vectors* of $A$,
- the factorization $USV^T$ is called the *singular value decomposition* of $A$.

# Singular Value Decomposition

$$A = USV^T,$$

$$U^T U = I, \quad V^T V = I,$$

- the columns of $U$ are orthonormal eigenvectors of $AA^T$,
- the columns of $V$ are orthonormal eigenvectors of $A^T A$,
- $S$ is a diagonal matrix containing the square roots of eigenvalues from $U$ or $V$ in descending order.

# Example

Let's find the singular value decomposition of $A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$.

In order to find $U$, we start with $AA^T$.

$$AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

- Eigenvalues: $\lambda_1 = 12$ and $\lambda_2 = 10$.
- Eigenvectors: $\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{u}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

# Example

Using Gram-Schmidt Process

$$\vec{v_1} = \vec{u_1} \quad \vec{w_1} = \frac{\vec{v_1}}{|\vec{v_1}|} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\vec{v_2} = \vec{u_2} - \frac{(\vec{u_2}, \vec{v_1})}{(\vec{v_1}, \vec{v_1})}\vec{v_1}$$

$$\vec{v_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - [0, 0] = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\vec{w_2} = \frac{\vec{v_2}}{|\vec{v_2}|} = \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]$$

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

# Example

The calculation of $V$ is similar. $V$ is based on $A^TA$, so we have

$$A^TA = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

We find the eigenvalues of $A^TA$

- Eigenvalues: $\lambda_1 = 12$, $\lambda_2 = 10$ and $\lambda_3 = 0$.
- Eigenvectors: $\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ and $\vec{u}_3 = \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$.

# Example

$$\vec{v_1} = \vec{u_1}, \quad \vec{w_1} = \frac{\vec{v_1}}{|v_1|} = \left[\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right]$$

$$\vec{v_2} = \vec{u_2} - \frac{(\vec{u_2}, \vec{v_1})}{(\vec{v_1}, \vec{v_1})}\vec{v_1} = [2, -1, 0]$$

$$\vec{w_2} = \frac{\vec{v_2}}{|\vec{v_2}|} = \left[\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0\right]$$

# Example

$$\vec{v}_3 = \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}\vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}\vec{v}_2 = \left[\frac{-2}{3}, \frac{-4}{3}, \frac{10}{3}\right]$$

$$\vec{w}_3 = \frac{\vec{v}_3}{|\vec{v}_3|} = \left[\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}}\right]$$

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix}, \quad V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix}$$

# Example

$$A = USV^T$$

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

# Principal Component Analysis

Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the eigenvalues of $A$ and $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n$ be a set of associated orthonormal eigenvectors.
Then the spectral decomposition of $A$ is given by

$$A = \lambda_1 \vec{x}_1 \vec{x}_1^T + \lambda_2 \vec{x}_2 \vec{x}_2^T + \cdots + \lambda_n \vec{x}_n \vec{x}_n^T.$$

If $A$ is a real $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, then an eigenvalue of largest magnitude is called *a dominant eigenvalue* of A.

# Principal Component Analysis

Let $X$ be the multivariate data matrix

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix}.$$

The measure of association between the $i$th and $k$th variables in the multivariate data matrix is given by the sample covariance

$$s_{ik} = \frac{1}{n} \sum_{j=1}^{n} \left(x_{ji} - \overline{x}_i\right) \left(x_{jk} - \overline{x}_k\right), \ \ i = 1, 2, \ldots, p, \ \ k = 1, 2, \ldots, p.$$

# Principal Component Analysis

- Let $S_n$ be the $p \times p$ covariance matrix associated with the multivariate data matrix $X$.

$$S_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

- Let the eigenvalues of $S_n$ be $\lambda_j, \quad j = 1, 2, \ldots, p$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and let the associated orthonormal eigenvectors be $\vec{u}_j, \quad j = 1, 2, \ldots, p$.
- The $i$th principal component $\vec{y}_i$ is given by the linear combination of the columns of $X$, where the coefficients are the entries of the eigenvector $\vec{u}_i$,

$$\vec{y}_i = X\vec{u}_i$$

# Principal Component Analysis

- The variance of $y_i$ is $\lambda_i$,
- The covariance of $y_i$ and $y_k, i \neq k$, is zero.
- If some of the eigenvalues are repeated, then the choices of the associated eigenvectors are not unique; hence the principal components are not unique.

$$\left( \begin{array}{l} \text{Proportion of the} \\ \text{total variance due} \\ \text{to the } k \text{ th principal} \\ \text{component} \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \; k = 1, 2, \ldots, p$$

# Example

Let's compute the first and second principal components for the multivariate data matrix $X$ given by

$$X = \begin{bmatrix} 39 & 21 \\ 59 & 28 \\ 18 & 10 \\ 21 & 13 \\ 14 & 13 \\ 22 & 10 \end{bmatrix}.$$

We first find the sample means $\vec{x}_1 \approx 28.8$ and $\vec{x}_2 \approx 15.8$. Then we take the matrix of sample means as

$$\vec{x} = \begin{bmatrix} 28.8 \\ 15.8 \end{bmatrix}$$

# Example

The variances are

$$s_{11} \approx 243.1 \quad \text{and} \quad s_{22} \approx 43.1$$

while the covariances are

$$s_{12} = s_{21} \approx 97.8.$$

We take the covariance matrix as

$$S_n = \begin{bmatrix} 243.1 & 97.8 \\ 97.8 & 43.1 \end{bmatrix}$$

# Example

- Eigenvalues: $\lambda_1 = 282.9744 \quad \text{and} \quad \lambda_2 = 3.2256$,
- Eigenvectors: $\vec{u}_1 = \begin{bmatrix} 0.9260 \\ 0.3775 \end{bmatrix}$ and $\vec{u}_2 = \begin{bmatrix} 0.3775 \\ -0.9260 \end{bmatrix}$.

  We find the first principal component as

$$\vec{y}_1 = 0.9260\,\mathrm{col}_1(X) + 0.3775\,\mathrm{col}_2(X).$$

  It follows that $\vec{y}_1$ accounts for the proportion

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} \quad (\text{ about } 98.9\%)$$

  of the total variance of $X$.

# Example

We find the second principal component as

$$\vec{y}_2 = 0.3775\,\mathrm{col}_1(X) - 0.9260\,\mathrm{col}_2(X).$$

It follows that $\vec{y}_2$ accounts for the proportion

$$\frac{\lambda_2}{\lambda_1 + \lambda_2} \quad (\text{ about } 0.011\%)$$

of the total variance of $X$.

# References


- Linear Algebra With Applications, 7th Edition
  by Steven J. Leon.
- Elementary Linear Algebra with Applications, 9th Edition
  by Bernard Kolman and David Hill.