More Linear Algebra

Yue Wu

Matrix diagonalization

decompose an n × n square matrix *A* into

$$A = V \Lambda V^{-1}$$

For example,

$$A = \begin{bmatrix} | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} | & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \\ | & | & | \end{bmatrix}^{-1}$$

square matrix

the inverse exists only if
eigenvectors are linearly independent

However, this is possible only if *A* is a square matrix and *A* has n linearly independent eigenvectors. Now, it is time to develop a solution for all matrices using SVD.

Singular vectors & singular values

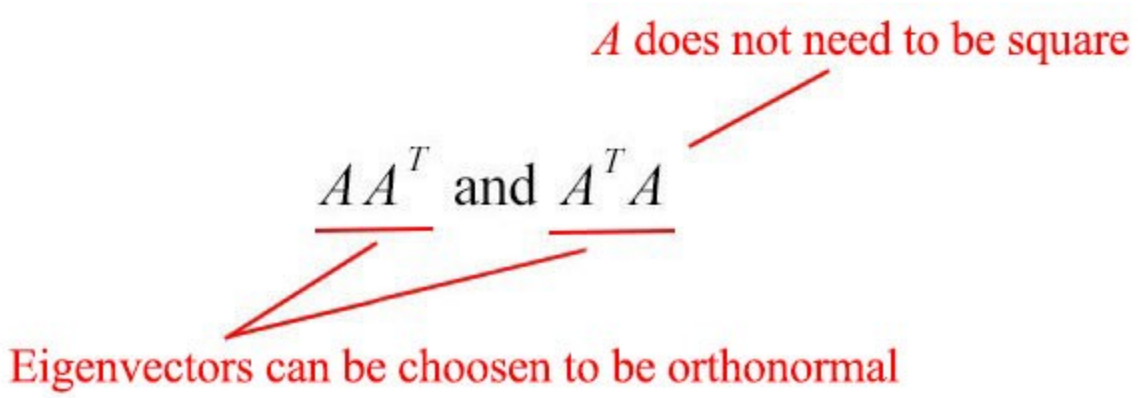
Consider any m × n matrix $\boldsymbol{A}$, we can multiply it with $\boldsymbol{A^T}$ to form $\boldsymbol{AA^T}$ and $\boldsymbol{A^TA}$ separately.

Claim:

- symmetric
- square
- at least positive semidefinite (eigenvalues are zero or positive)
- both matrices have the same positive eigenvalues
- both have the same rank *r* as *A*

Symmetric Matrices

The covariance matrices that we often use in ML are in this form. Since they are symmetric, we can choose its eigenvectors to be **orthonormal** (perpendicular to each other with unit length) — this is a fundamental property for symmetric matrices.



Defns

We name the eigenvectors for $\boldsymbol{AA^T}$ as u_i and $\boldsymbol{A^TA}$ as v_i here and call these sets of eigenvectors u and v the **singular vectors** of A. Both matrices have the same positive eigenvalues. The square roots of these eigenvalues are called **singular values**.

Defns

$$U \qquad\qquad V$$

$$\begin{pmatrix} | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ | & & | \end{pmatrix} \qquad \begin{pmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{pmatrix}$$

$$\begin{pmatrix} u_{11} & & u_{m1} \\ & \ddots & \\ u_{1m} & & u_{mm} \end{pmatrix}$$

Observation (for Orthonormal Matrices)

$$U^T U = I$$

$$V^T V = I$$

SVD

- SVD states that **any** matrix *A* can be factorized as:

$$A = U S V^T$$

where *U* and *V* are orthogonal matrices with orthonormal eigenvectors chosen from $\boldsymbol{AA^T}$ and $\boldsymbol{A^TA}$ respectively. $\boldsymbol{S}$ is a diagonal matrix with *r* elements equal to the square root of the positive eigenvalues of AA^T or $A^T A$ (both matrices have the same positive eigenvalues anyway). The diagonal elements are composed of singular values.

$$S$$

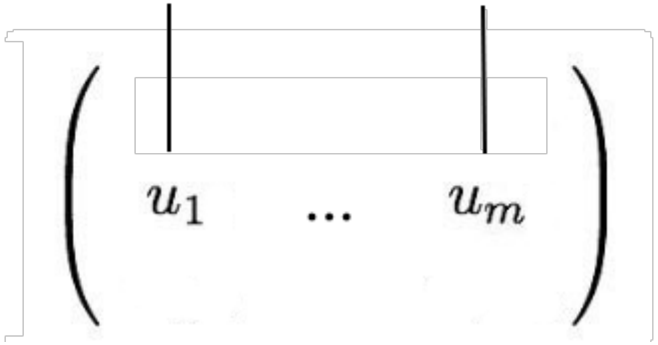
$$\begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & \\ & & \ddots & & \\ & & & \sqrt{\lambda_r} & \\ & & & & \ddots \; 0 \end{pmatrix} \equiv \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \ddots \; 0 \end{pmatrix}$$

σ_2: *singular value*

$$\underset{m \times n}{\overset{\mathbf{A}}{\begin{pmatrix} x_{11} & x_{12} & & x_{1n} \\ & & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} = \underset{m \times m}{\overset{\mathbf{U}}{\begin{pmatrix} u_{11} & & u_{m1} \\ & \ddots & \\ u_{1m} & & u_{mm} \end{pmatrix}}} \underset{m \times n}{\overset{\mathbf{S}}{\begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ & & & \ddots \\ 0 & & & 0 \end{pmatrix}}} \underset{n \times n}{\overset{\mathbf{V}^{\mathrm{T}}}{\begin{pmatrix} v_{11} & & v_{1n} \\ & \ddots & \\ v_{n1} & & v_{nn} \end{pmatrix}}}$$

We can arrange eigenvectors in different orders to produce *U* and *V*. To standardize the solution, we order the eigenvectors such that vectors with higher eigenvalues come before those with smaller values.

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m$$



Example

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}, \qquad A^T A = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

Example continued

$$AA^T = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

eigenvalues: $\lambda_1 = 25$, $\lambda_2 = 9$

eigenvectors

$$u_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad u_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

$$A^TA = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

eigenvalues: $\lambda_1 = 25$, $\lambda_2 = 9$, $\lambda_3 = 0$

eigenvectors

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix} \quad v_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ -1/3 \end{pmatrix}$$

$$A = USV^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \\ 2/3 & -2/3 & -1/3 \end{pmatrix}$$

Covariance matrices

Variance measures how a variable varies between itself while covariance is between two variables (*a* and *b*).

$$\sigma_{ab}{}^2 = \text{cov}(a, b) = \text{E}[(\ a - \overline{a}\ )(\ b - \overline{b}\ )]$$

$$\sigma_a{}^2 = \text{var}(a) = \text{cov}(a, a) = \text{E}[(\ a - \overline{a}\ )^2]$$

We can hold all these possible combinations of covariance in a matrix called the **covariance matrix** Σ.

$$\Sigma = \begin{pmatrix} E[(x_1 - \mu_1)(x_1 - \mu_1)] & E[(x_1 - \mu_1)(x_2 - \mu_2)] & \dots & E[(x_1 - \mu_1)(x_p - \mu_p)] \\ E[(x_2 - \mu_2)(x_1 - \mu_1)] & E[(x_2 - \mu_2)(x_2 - \mu_2)] & \dots & E[(x_2 - \mu_2)(x_p - \mu_p)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_p - \mu_p)(x_1 - \mu_1)] & E[(x_p - \mu_p)(x_2 - \mu_2)] & \dots & E[(x_p - \mu_p)(x_p - \mu_p)] \end{pmatrix}$$

$$\Sigma = \mathrm{E}[(X - \bar{X})(X - \bar{X})^{\mathrm{T}}]$$

$$\Sigma = \frac{XX^{\mathrm{T}}}{n} \quad \text{(if } X \text{ is already zero centered)}$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

Unit vector.

Each column is a sample.

$$\vec{v}^T X = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{v}^T \vec{x}_1 & \vec{v}^T \vec{x}_2 & \vec{v}^T \vec{x}_3 & \vec{v}^T \vec{x}_4 \end{bmatrix}$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$\text{Variance of Projected Data} = \vec{v}^T X X^T \vec{v}$$

Want to maximize this subject to $\vec{v}$ being a unit vector.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

↑

Lagrange multipliers: next week!

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

PCA

Find vector $\vec{v}$ such that variance of projected data is maximized.

$$L(\vec{v}) = \left(\vec{v}^T X X^T \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\right)$$

$$\frac{\partial L}{\partial \vec{v}} = 2XX^T\vec{v} - 2\lambda\vec{v}$$

$$0 = (XX^T - \lambda I)\vec{v}$$

Thus, we are just looking for the eigenvector of this matrix.

Can also think of the equivalent SVD problem.