Lecture 14, Oct 25, 2023

Eigendecomposition

Definition

An *eigenvector* $\boldsymbol{x} \neq \boldsymbol{0}$ of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is any vector satisfying

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

for some *eigenvalue* $\lambda \in \mathbb{C}$.

Definition

The *spectrum* of $\boldsymbol{A}$ is the set of all eigenvalues of $\boldsymbol{A}$.

The *spectral radius* of $\boldsymbol{A}$ is $\rho(\boldsymbol{A}) = \max|\lambda|$ overall eigenvalues λ of $\boldsymbol{A}$.

- Eigenvalues are connected to optimization; consider the optimization: $\min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ subject to $\|\boldsymbol{x}\|_2 = 1$
  - This is a quadratically constrained quadratic program (QCQP)
  - The Lagrangian is $\Lambda(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} - \lambda(\boldsymbol{x}^T \boldsymbol{x} - 1)$
  - From $\dfrac{\partial \Lambda}{\partial \boldsymbol{x}^T} = 0$ we get that $2\boldsymbol{A}\boldsymbol{x} - 2\lambda \boldsymbol{x} = 0 \implies \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$
  - Therefore the critical points are the unit eigenvectors of $\boldsymbol{A}$ (unit due to the constraint)
- Important properties:
  - Eigenvectors corresponding to different eigenvalues are linearly independent
  - $\boldsymbol{A}$ is diagonalizable if its eigenvectors span $\mathbb{R}^n$, in which case $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \iff \boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}$, which is called a *similarity transformation*
    - If $\boldsymbol{A}$ is not diagonalizable, it is *degenerate* or has *degenerate eigenvalues*
    - If $\boldsymbol{V}$ is orthogonal, we can interpret this as decomposing $\boldsymbol{A}$ into a pure rotation, a pure scaling, and then the inverse of the pure rotation
  - *Spectral theorem*: Symmetric matrices have n orthonormal eigenvectors with (possibly repeated) real eigenvalues
  - Positive definite matrices have all positive eigenvalues; positive semi-definite matrices have all nonnegative eigenvalues

Singular Value Decomposition

Definition

The *singular value decomposition* (SVD) of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is

$$\boldsymbol{U}^T \boldsymbol{A}\boldsymbol{V} = \boldsymbol{\Sigma} \iff \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

where $\boldsymbol{V} \in \mathbb{R}^{n \times n}, \boldsymbol{U} \in \mathbb{R}^{m \times m}$ are orthogonal matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix which contains the singular values of $\boldsymbol{A}$.

Note that $\boldsymbol{V}$ contains the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{U}$ contains the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$, while the $\sigma_i \in \mathbb{R}$ singular values on the diagonal of $\boldsymbol{\Sigma}$ are the square roots of the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$.

- Note $\boldsymbol{\Sigma}$ has the same dimension as $\boldsymbol{A}$
- $\boldsymbol{A}^T\boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{A}^T$ are symmetric, so by the spectral theorem their eigenvalues are orthonormal; furthermore, $\boldsymbol{A}^T\boldsymbol{A}$ is positive semi-definite so eigenvalues are all nonnegative, thus the singular values are real

- $\boldsymbol{\Sigma}$ has the singular values $\sigma_1, \dots, \sigma_m$ (assuming $m \leq n$) on its diagonal; note that it can be rectangular
  - We can split up $\boldsymbol{\Sigma}$ into a square diagonal matrix and a block of all zeroes
  - In general $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & & \mathbf{0} \\ & \ddots & & \vdots \\ & & \sigma_k & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}$
  - By convention the singular values are sorted such that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq \sigma_{k+1} = \sigma_{k+2} = \dots = 0$
  - The first $k = \operatorname{rank} \boldsymbol{A}$ singular values are nonzero (note k is also the number of strictly positive eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$)
- Note that $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma} \implies \boldsymbol{A}\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$; therefore we see that the SVD is a generalization of eigendecomposition for non-square $\boldsymbol{A}$
  - $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k & \boldsymbol{v}_{k+1} & \boldsymbol{v}_{k+2} & \cdots & \boldsymbol{v}_n \end{bmatrix}$
    * $\boldsymbol{v}_1$ to $\boldsymbol{v}_k$ are the normalized eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$
    * $\boldsymbol{v}_{k+1}$ to $\boldsymbol{v}_n$ are taken from the null space of $\boldsymbol{A}^T\boldsymbol{A}$
    * All $\boldsymbol{v}_j$ are chosen such that $\boldsymbol{V}$ is orthogonal
  - $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_k & \boldsymbol{u}_{k+1} & \boldsymbol{u}_{k+2} & \cdots & \boldsymbol{u}_n \end{bmatrix}$
    * $\boldsymbol{u}_1$ to $\boldsymbol{u}_k$ are the normalized eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$
    * $\boldsymbol{u}_{k+1}$ to $\boldsymbol{u}_n$ are taken from the null space of $\boldsymbol{A}\boldsymbol{A}^T$
    * All $\boldsymbol{u}_j$ are chosen such that $\boldsymbol{U}$ is orthogonal
    * This is because $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T(\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{U}^T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ which is just a diagonalization; ditto for $\boldsymbol{V}$
- Example: find the SVD of $\begin{bmatrix} 1 & 0 \end{bmatrix}$
  - $\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
    * This has rank 1, giving eigenvalues $\lambda_1 = 1, \lambda_2 = 0$
    * We get the nonzero singular value $\sigma_1 = \sqrt{\lambda_1} = 1$
    * The unit eigenvector corresponding to λ_1 is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$
    * To get the second vector in $\boldsymbol{V}$ we take $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which is orthogonal to the first and is a zero eigenvector
  - To find $\boldsymbol{u}_j$ we can use $\boldsymbol{A}\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$, giving $u_1 = 1$
  - Therefore $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \boldsymbol{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{U} = \begin{bmatrix} 1 \end{bmatrix}$
- Example: find the SVD of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$
  - $\boldsymbol{A}^T\boldsymbol{A} = 1$ giving $\lambda_1 = 1 \implies \sigma_1 = 1$ and an eigenvector of 1
  - $\boldsymbol{A}\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j \implies \boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} 1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
    * Choose $\boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ so that it is orthogonal to $\boldsymbol{u}_1$ and gives $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{u}_2 = \mathbf{0}$
  - Therefore $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{V} = \begin{bmatrix} 1 \end{bmatrix}, \boldsymbol{U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
  - We could have found this using the previous example's result, since the matrix was just transposed
- If $\boldsymbol{A}_1 = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^T$ and $\boldsymbol{A}_1^T = \boldsymbol{A}_2 = \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^T$, then $\boldsymbol{V}_1 = \boldsymbol{U}_2, \boldsymbol{U}_1 = \boldsymbol{V}_2, \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2^T$
- Consider the effect of $\boldsymbol{A}$ on length: examine the critical points of $\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}$ (the *Rayleigh quotient*) subject to $\|\boldsymbol{x}\|_2 = 1$
  - From the example in the previous section we know that critical points are the unit eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$: $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$, where all eigenvalues are nonnegative and eigenvectors are orthonormal
  - Let $\tilde{\boldsymbol{u}}_i = \boldsymbol{A}\boldsymbol{v}_i$, then $\lambda_i \tilde{\boldsymbol{u}}_i = \lambda_i \boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{A}(\lambda_i \boldsymbol{v}_i) = \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{A}\boldsymbol{A}^T\tilde{\boldsymbol{u}}_i$ so $\tilde{\boldsymbol{u}}_i$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}^T$ with the same eigenvalue λ_i
    * $\|\tilde{\boldsymbol{u}}_i\|_2 = \sqrt{\boldsymbol{v}_i^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i} = \sqrt{\lambda_i \boldsymbol{v}_i^T\boldsymbol{v}_i} = \sqrt{\lambda}\|\boldsymbol{v}_i\|_2 = \sqrt{\lambda}$

    - * We can normalize to get $\boldsymbol{u}_i = \frac{\tilde{\boldsymbol{u}}_i}{\lambda_i}$
  - So we get two eigenproblems $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i, \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i$ with all $\boldsymbol{v}_i, \boldsymbol{u}_1$ being orthonormal
  - Now define $\boldsymbol{U}_1 = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k\end{bmatrix}, \boldsymbol{V}_1 = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k\end{bmatrix}$ and we can show that this leads us to the SVD
- $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T\end{bmatrix}$ where $\boldsymbol{U}_2$ and $\boldsymbol{V}_2$ are chosen in the null spaces of $\boldsymbol{A}\boldsymbol{A}^T$, $\boldsymbol{A}^T\boldsymbol{A}$ to give orthogonality
  - Essentially, $\boldsymbol{A}$ takes everything from span$\{\boldsymbol{V}_1\}$ and maps it to span$\{\boldsymbol{U}_1\}$; everything in span$\{\boldsymbol{V}_2\}$ is mapped to zero, so everything in span$\{\boldsymbol{U}_2\}$ are all the points that $\boldsymbol{A}$ cannot reach
  - $\boldsymbol{V}_1, \boldsymbol{U}_1$ are unique, but $\boldsymbol{V}_2, \boldsymbol{U}_2$ can be chosen arbitrarily as long as orthogonality is maintained
  - In practice, we rarely need to compute $\boldsymbol{V}_2, \boldsymbol{U}_2$
- The SVD allows us to write $\boldsymbol{A}$ using its modes: $\boldsymbol{A} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$
  - Since the σ_i are sorted, the earlier terms in the summation are more "important"
  - We can cut of this summation to get a version of $\boldsymbol{A}$ with some of the unimportant bits removed