Modified on Monday 2nd November, 2020, 12:49:4812:49

Nov 2, 2020

- We will begin with some review of what we have done before midterm break.
- Recall that the last subject was SVD. Any matrix $A$ of rank $k$ has an SVD,

$$A = U\Sigma V^T = U \begin{bmatrix} \Sigma_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^T = U_k\Sigma_k V_k^T = \sigma_1 \mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T.$$

**Reduced SVD.** The matrix $A$ can be expressed as $A = U_k\Sigma_k V_k^T$ where $U_k$ is an $m \times k$ with orthonormal columns and $V_k$ is an $n \times k$ with orthonormal columns and $\Sigma_k = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$ is a $k \times k$ diagonal matrix with positive diagonal entries, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$. Let $U_k = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{bmatrix}$ and $V_k = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix}$. The values $\sigma_1, \sigma_2, \ldots, \sigma_k$ are called singular values of $A$. Then

$$A = \sigma_1 \mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T.$$

The right singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are eigenvectors of $A^TA$ corresponding to eigenvalues $\sigma_1^2, \ldots, \sigma_k^2$, respectively. The left singular vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are eigenvectors of $AA^T$ corresponding to eigenvalues $\sigma_1^2, \ldots, \sigma_k^2$, respectively. The right and left singular vectors are related by $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ and $A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i$.

- Assume that $A$ has rank $r$. For $k \leq r$, define

$$A_k = \sigma_1 \mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T.$$

We also have a theorem regarding the approximation by a low rank matrix.

Theorem. (Eckart-Young) Let $A$ and $A_k$ be matrices defined in the above. If $B$ has rank $k$, then $\|A - B\| \geq \|A - A_k\|$.

The above theorem holds for the following matrix norms.

  - Spectral norm: $\|A\|_2 = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \sigma_1$
  - Frobenius norm: $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} = \sqrt{\operatorname{tr}(A^TA)}$
  - Nuclear norm: $\|A\|_N = \sigma_1 + \cdots + \sigma_r$

- Let $A = (a_{ij}) = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ be $2 \times n$ matrix with $\sum_{j=1}^n a_{ij} = 0$ for $i = 1$ and 2. We may regard $A$ as a centered data. Find the direction $\mathbf{x}_1$ of the line such that the sum of squared distances from the data points $(a_{1j}, a_{2j})$ to the line is a minimum.

Assume that $A$ has rank 2. Let $A = U\Sigma V^T$ be the (reduced) singular value decomposition of $A$ with $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$, $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2)$ and $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$.

$$A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T$$

Choose an orthonormal basis for $\mathbb{R}^2$, $\{\mathbf{x}_1, \mathbf{x}_2\}$. We want to minimize $\sum_{j=1}^n |\mathbf{a}_j^T\mathbf{x}_2|^2$.

$$\begin{aligned} \|A\|_F^2 = \sigma_1^2 + \sigma_2^2 = \operatorname{tr}(A^TA) = \sum_{j=1}^n \|\mathbf{a}_j\|^2 &= \sum_{j=1}^n |\mathbf{a}_j^T\mathbf{x}_1|^2 + \sum_{j=1}^n |\mathbf{a}_j^T\mathbf{x}_2|^2 \\ &= \sum_{j=1}^n \mathbf{x}_1^T\mathbf{a}_j\mathbf{a}_j^T\mathbf{x}_1 + \sum_{j=1}^n \mathbf{x}_2^T\mathbf{a}_j\mathbf{a}_j^T\mathbf{x}_2 \\ &= \mathbf{x}_1^TAA^T\mathbf{x}_1 + \mathbf{x}_2^TAA^T\mathbf{x}_2 \end{aligned}$$

We can minimize the second summand by maximizing the first summand, i.e., by choosing $\mathbf{x}_1 = \mathbf{u}_1$ and $\mathbf{x}_2 = \mathbf{u}_2$. $\|A - A_1\|_F^2 = \sigma_2^2$ is minimal. Note that

$$A^T\mathbf{u}_1 = \sigma_1\mathbf{v}_1, \quad A_1 = \sigma_1\mathbf{u}_1\mathbf{v}_1^T.$$

Project data points along $\mathbf{u}_1$ to obtain

$$\mathbf{u}_1(\mathbf{u}_1^T A) = \sigma_1\mathbf{u}_1\mathbf{v}_1^T = A_1.$$

- Find the eigenvalues of $S$. Recall that we found the largest eigenvalue by $\max_{\|\mathbf{x}\|=1}\mathbf{x}^T S\mathbf{x}$.
- Rayleigh quotients and generalized eigenvalues
- Rayleigh quotient. Let $S$ be a symmetric matrix. The *Rayleigh quotient* $R(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$ is defined as

$$R(\mathbf{x}) = \frac{\mathbf{x}^T S\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

  Note that $R(k\mathbf{x}) = R(\mathbf{x})$ for all nonzero $k \in \mathbb{R}$.
- Generalized Rayleigh quotient. Let $S$ and $M$ be $n \times n$ symmetric matrix. The quotient

$$R(\mathbf{x}) = \frac{\mathbf{x}^T S\mathbf{x}}{\mathbf{x}^T M\mathbf{x}}, \quad \mathbf{x} \neq \mathbf{0},$$

  is called the *generalized Rayleigh quotient*.
- Maximize $R(\mathbf{x})$. This can be done easily, if $M$ is positive definite. Assume that $M$ is positive definite. Let $H = M^{-1/2}SM^{-1/2}$, $\mathbf{y} = M^{-1/2}\mathbf{x}$.

$$\max R(\mathbf{x}) = \frac{\mathbf{x}^T S\mathbf{x}}{\mathbf{x}^T M\mathbf{x}} = \max\frac{\mathbf{y}^T H\mathbf{y}}{\mathbf{y}^T\mathbf{y}}$$

  This value is the largest eigenvalue of $H$, equivalently, $M^{-1}S$.
- Generalized eigenvalue.

  Let $S$ and $M$ be symmetric matrices. If $S\mathbf{x} = \lambda_1 M\mathbf{x}$ and $S\mathbf{y} = \lambda_2 M\mathbf{y}$ and $\lambda_1 \neq \lambda_2$ then $\mathbf{x}^T M\mathbf{y} = 0$.

  Proof.

$$\mathbf{y}^T S\mathbf{x} = \lambda_1\mathbf{y}^T M\mathbf{x}, \quad \mathbf{x}^T S\mathbf{y} = \lambda_2\mathbf{x}^T M\mathbf{y}$$

  Since these two are the same, $\mathbf{x}^T M\mathbf{y} = 0$.