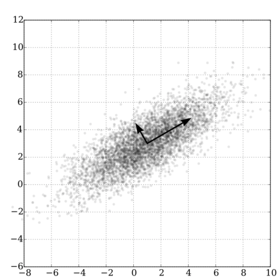
Modified on Wednesday 28th October, 2020, 12:50:5812:50

Oct 28, 2020

- Principal components and the best low rank matrix
- Recall singular value decomposition. Any $n \times n$ matrix A can be expressed as $A = U\Sigma V^T$ where U and V are $n \times n$ orthogonal matrices and Σ is a diagonal matrix with nonnegative diagonal entries.
- Singular value decomposition and reduced singular value decomposition.
  Let A be an $m \times n$ matrix of rank k.
  - **SVD.** The matrix A can be expressed as $A = U\Sigma V^T$ where U is an $m \times m$ orthogonal matrix and V is an $n \times n$ orthogonal matrix and $\Sigma = (\sigma_{ij})$ is an $m \times n$ matrix satisfying $\sigma_{i,j} = 0$ for all $i \neq j$, $\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{kk} > 0$ and $\sigma_{ii} = 0$ for $i > k$. Let $U = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix}$ and $V = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ and $\sigma_i = \sigma_{ii}$. 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 $A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i$.
  - **Reduced SVD.** The matrix A can be expressed as $A = U_k\Sigma_kV_k^T$ 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 $A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i$.
  - Any $m \times n$ matrix A of rank k can be expressed as
    $$A = U\Sigma V^T = U\begin{bmatrix}\Sigma_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}V^T = U_k\Sigma_kV_k^T = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^T.$$
- What is an SVD of I_n? What is an SVD of an orthogonal matrix Q? Is an SVD unique?
- Note that $\mathbf{u}_i\mathbf{v}_i^T$ is an $m \times n$ matrix of rank one.
- Principal component analysis (PCA): Given data samples $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$, we want to find the k-dimensional projection of $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ that preserves most of the variance of the $\mathbf{x}_i$.



- The sample mean vector and sample covariance matrix are defined by

$$\mu = \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i, \text{ and } S = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T = \frac{1}{m-1}AA^T, \text{ where } A = \begin{bmatrix} \mathbf{x}_1 - \mu & \cdots & \mathbf{x}_m - \mu \end{bmatrix}.$$

- In order to maximize variance, find $\mathbf{u}_1$ such that

$$\mathbf{u}_1 = \arg\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top AA^T\mathbf{u}.$$

  This is an eigenvector of AA^T associated with the largest eigenvalue of AA^T. This corresponds to left singular vector of A.
- Given a data matrix A, analyze the data by approximating A with the closest matrix of rank k, for some positive integer k.
  - Find the closest rank k matrix to A.
  - How do we measure the distance between matrices?
- Let A be an $m \times n$ matrix of rank r with SVD,

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

  For $k \le r$, define

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

- Theorem. (Eckart-Young) Let A and A_k be matrices defined in the above. If B has rank k, then $\|A - B\| \ge \|A - A_k\|$.
- The above theorem holds for the following matrix norms.
  - Spectral norm: $\|A\|_2 = \max \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sigma_1$
  - Frobenius norm: $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} = \sqrt{\mathrm{tr}(A^TA)}$
  - Nuclear norm: $\|A\|_N = \sigma_1 + \cdots + \sigma_r$

  Compute above norms for I_n and an $n \times n$ orthogonal matrix Q.

$$\|Q\|_2 = 1, \quad \|Q\|_F = \sqrt{n}, \quad \|Q\|_N = n$$

- Theorem. (Eckart-Young) If B has rank k, then $\|A - B\|_2 \ge \|A - A_k\|_2$. If $\mathrm{rank}(B) \le k$ then

$$\|A - B\|_2 = \max \frac{\|(A - B)\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \ge \sigma_{k+1}.$$

- Theorem. (Eckart-Young) If B has rank k, then $\|A - B\|_F \ge \|A - A_k\|_F$. If $\mathrm{rank}(B) \le k$ then

$$\|A - B\|_F \ge \sqrt{\sigma_{k+1} + \cdots + \sigma_r}.$$

- Least squares and perpendicular least squares (total least squares, orthogonal regression)