Modified on Monday 26$^{\text{th}}$ October, 2020, 11:34:1311:34

Oct 26, 2020

- Singular values and singular vectors in SVD. [8.6]
- Every orthogonal set of nonzero vectors in $\mathbb{R}^n$ can be extended to an orthogonal basis for $\mathbb{R}^n$.
- Theorem. Every nonzero subspace of $\mathbb{R}^n$ has an orthonormal basis.

  Proof by construction. Let W be a nonzero subspace of $\mathbb{R}^n$ and $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ an ordered basis for W. We construct an ordered orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ for W as follows.

  - Let
  $$\mathbf{v}_1 = \mathbf{w}_1, \quad \mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$$
  - Assume that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_j\}$ has been constructed with $j < k$. Let $W_j = \operatorname{span}(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j) = \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_j)$. Define
  $$\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \operatorname{proj}_{W_j}\mathbf{w}_{j+1} = \mathbf{w}_{j+1} - \sum_{i=1}^{j} \frac{\mathbf{w}_{j+1}\cdot\mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i = \mathbf{w}_{j+1} - \sum_{i=1}^{j}(\mathbf{w}_{j+1}\cdot\mathbf{q}_i)\mathbf{q}_i, \quad \mathbf{q}_{j+1} = \frac{\mathbf{v}_{j+1}}{\|\mathbf{v}_{j+1}\|}$$
  - Repeating the above step, we can construct $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$, which is an orthonormal basis for W.

- From the relations
  $$\mathbf{v}_1 = \mathbf{w}_1, \quad \mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|},$$
  and
  $$\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \sum_{i=1}^{j}(\mathbf{w}_{j+1}\cdot\mathbf{q}_i)\mathbf{q}_i, \quad \mathbf{q}_{j+1} = \frac{\mathbf{v}_{j+1}}{\|\mathbf{v}_{j+1}\|}$$
  we obtain
  $$\mathbf{w}_1 = (\mathbf{w}_1\cdot\mathbf{q}_1)\mathbf{q}_1, \quad \mathbf{w}_{j+1} = \sum_{i=1}^{j+1}(\mathbf{w}_{j+1}\cdot\mathbf{q}_i)\mathbf{q}_i.$$
  This can be written as a matrix decomposition. Let $A = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_k\end{bmatrix}$ and $Q = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k\end{bmatrix}$. Then
  $$A = Q\begin{bmatrix}\mathbf{w}_1\cdot\mathbf{q}_1 & \mathbf{w}_2\cdot\mathbf{q}_1 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_1 \\ 0 & \mathbf{w}_2\cdot\mathbf{q}_2 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_k\end{bmatrix}.$$
  $$\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_k\end{bmatrix} = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k\end{bmatrix}\begin{bmatrix}\mathbf{w}_1\cdot\mathbf{q}_1 & \mathbf{w}_2\cdot\mathbf{q}_1 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_1 \\ 0 & \mathbf{w}_2\cdot\mathbf{q}_2 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{w}_k\cdot\mathbf{q}_k\end{bmatrix}$$
  Let R denote the right most matrix in the above. Then we have $A = QR$, where Q is an $n \times k$ matrix whose columns are orthonormal and R is an invertible upper triangular matrix. The decomposition of a matrix A with full column rank into the product QR is called a **QR-decomposition of A**. Since $Q^TQ = I_k$, we have
  $$R = Q^TA,$$
  which implies that a QR-decomposition of A can be computed by the Gram-Schmidt process.

- If A has a QR-decomposition, then the normal equation associated with $A\mathbf{x} = \mathbf{b}$,
$$A^T A\mathbf{x} = A^T\mathbf{b},$$
changes into
$$(QR)^T(QR)\mathbf{x} = (QR)^T\mathbf{b}, \quad R^T R\mathbf{x} = R^T Q^T\mathbf{b}, \quad R\mathbf{x} = Q^T\mathbf{b}.$$
So the least squares solution is
$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$
- Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the reflection in $\mathbb{R}^n$ about a hyperplane $\mathbf{a}^\perp$.
$$T(\mathbf{x}) = \mathbf{x} - 2\mathrm{proj}_{\mathbf{a}}\mathbf{x} = \mathbf{x} - 2\,\frac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\,\mathbf{a}.$$
- Singular value decomposition
- Recall that real symmetric matrices are orthogonally diagonalizable; for any $n\times n$ symmetric matrix A, there exist an orthogonal matrix P and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ satisfying $A = PDP^T$, equivalently, $P^T AP = D$.
- Theorem. Let A be an $n \times n$ matrix of rank k. Then A can be factored as
$$A = U\Sigma V^T$$
where U and V are $n \times n$ orthogonal matrices and Σ is an $n \times n$ diagonal matrix
$$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0)$$
with $\sigma_i > 0$ for all $i = 1, \dots, k$.

  Proof. $A^T A$ is symmetric, of rank k.

  For some orthogonal matrix $V = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ and diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$,
$$A^T A = VDV^T.$$
  Eigenvalues of $A^T A$ are nonnegative: If $\mathbf{x}$ is an eigenvector, then $\|A\mathbf{x}\|^2 = A\mathbf{x}\cdot A\mathbf{x} = \lambda\|\mathbf{x}\|^2$.

  Rearrange the columns of V so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Since $A^T A$ and D are similar, they have the same rank, i.e. k. So we have $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ and $\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0$.

  The set $\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\}$ is an orthogonal set of nonzero vectors: $A\mathbf{v}_i\cdot A\mathbf{v}_j = \mathbf{v}_i\cdot A^T A\mathbf{v}_j = \lambda_j(\mathbf{v}_i\cdot\mathbf{v}_j) = 0$. $\|A\mathbf{v}_i\|^2 = \lambda_i$. Normalzing the vectors as
$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}, \quad i = 1, \dots, k$$
  yield an orthonormal set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. Extend it to an orthonormal basis for $\mathbb{R}^n$,
$$\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}.$$
  Let $U = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}$ and $\Sigma = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_k}, 0, \dots, 0)$. Then
$$\begin{aligned} U\Sigma &= \begin{bmatrix}\sqrt{\lambda_1}\mathbf{u}_1 & \cdots & \sqrt{\lambda_k}\mathbf{u}_k & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix} \\ &= \begin{bmatrix}A\mathbf{v}_1 & \cdots & A\mathbf{v}_k & A\mathbf{v}_{k+1} & \cdots & A\mathbf{v}_n\end{bmatrix} \\ &= AV \end{aligned}$$
  The positive values $\sigma_i = \sqrt{\lambda_i}$, $1 \leq i \leq k$, are called **singular values** of A.
- The singular decomposition of A
$$A = U\Sigma V^T = U\mathrm{diag}(\sigma_1, \dots, \sigma_k, 0, \dots, 0)V^T$$