Modified on Wednesday 14th October, 2020, 10:45:0910:45

Oct 12, 2020

- Projection, Gram-Schmidt process, QR-decomposition. [7.7, 7.8, 7.9, 7.10]
- Recall: Orthogonal projection of $\mathbf{x}$ in $\mathbb{R}^n$ onto the line through the origin making an angle θ with the positive x-axis can be expressed as

$$P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$$

- Orthogonal projection of $\mathbf{x}$ along $\mathbf{a}$ (or onto the line $W = \text{span}\{\mathbf{a}\}$)

$$\text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\mathbf{a} = \frac{\mathbf{a}\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}$$

- Theorem. Let W be a subspace of $\mathbb{R}^n$. For any $\mathbf{x} \in \mathbb{R}^n$, we can express $\mathbf{x}$ as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ for some $\mathbf{x}_1 \in W$ and $\mathbf{x}_2 \in W^\perp$. Moreover, the expression is unique.

  Proof. Choose an $n \times k$ matrix A whose columns form a basis for W. $W = \text{col}(A)$. A^TA is invertible.

  Assume that $\mathbf{x}_1 = A\mathbf{v}$. Then $\mathbf{x}_2 = \mathbf{x} - A\mathbf{v}$.

$$A^T\mathbf{x}_2 = A^T(\mathbf{x} - A\mathbf{v}) = \mathbf{0}, \quad A^TA\mathbf{v} = A^T\mathbf{x}, \quad \mathbf{v} = (A^TA)^{-1}A^T\mathbf{x}$$

  Let $\mathbf{x}_1 = A\mathbf{v} = A(A^TA)^{-1}A^T\mathbf{x}$, and $\mathbf{x}_2 = \mathbf{x} - \mathbf{x_1}$.

- The **orthogonal projection of $\mathbf{x}$ on W**:

$$\text{proj}_W\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{x}$$

$$\mathbf{x} = \text{proj}_W\mathbf{x} + \text{proj}_{W^\perp}\mathbf{x}$$

- Best approximation and least squares
- Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$. Can we find a vector $\hat{\mathbf{w}} \in W$ such that $\|\mathbf{x} - \hat{\mathbf{w}}\| < \|\mathbf{x} - \mathbf{w}\|$ for all $\mathbf{w} \in W$ different from $\hat{\mathbf{w}}$. The vector $\hat{\mathbf{w}}$, if it exists, is called a **best approximation to $\mathbf{x}$ from W**.
- Theorem. Let W be a subspace of $\mathbb{R}^n$. For any $\mathbf{x} \in \mathbb{R}^n$, $\text{proj}_W\mathbf{x}$ is a best approximation to $\mathbf{x}$ from W.

  Proof. For any $\mathbf{w} \in W$,

$$\mathbf{x} - \mathbf{w} = (\mathbf{x} - \text{proj}_W\mathbf{x}) + (\text{proj}_W\mathbf{x} - \mathbf{w}).$$

  Since $\mathbf{x} - \text{proj}_W\mathbf{x}$ is in $W^\perp$ and $\text{proj}_W\mathbf{x} - \mathbf{w}$ is in W,

$$\|\mathbf{x} - \mathbf{w}\|^2 = \|\mathbf{x} - \text{proj}_W\mathbf{x}\|^2 + \|\text{proj}_W\mathbf{x} - \mathbf{w}\|^2.$$

  So, if $\mathbf{w} \neq \text{proj}_W\mathbf{x}$, then

$$\|\mathbf{x} - \text{proj}_W\mathbf{x}\| < \|\mathbf{x} - \mathbf{w}\|.$$

- The distance from $\mathbf{x}$ to W:

$$\|\mathbf{x} - \text{proj}_W\mathbf{x}\| = \|\text{proj}_{W^\perp}\mathbf{x}\|.$$

- Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. A vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is called a best approximate solution or a **least squares solution** of $A\mathbf{x} = \mathbf{b}$ if

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

  for all $\mathbf{x} \in \mathbb{R}^n$.

- Least squares solution, least squares error vector, least squares error
- **Normal equation associated with $A\mathbf{x} = \mathbf{b}$**:

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

- The least squares solutions to $A\mathbf{x} = \mathbf{b}$ are the solutions to the assorted normal equation

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

If A has full column rank, then the normal equation has a unique solution $\hat{\mathbf{x}}$,

$$\hat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}.$$

- Given a set of data points, (x_i, y_i) for $i = 1, \ldots, n$, find a polynomial of degree d

$$y(x) = a_0 + a_1 x + \cdots + a_d x^d$$

which fits the data as closely as possible.

What do we mean by *as closely as possible*?

  - Minimize $\sum_{i=0}^{d} |y_i - y(x_i)|$
  - Minimize $\sum_{i=0}^{d} |y_i - y(x_i)|^2$

Do they give the same result? Which is better?

Consider the linear system $M\mathbf{a} = \mathbf{y}$ with $M = (x_i^j)$, $\mathbf{a} = (a_i)$ and $\mathbf{y} = (y_i)$.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Solve for $\mathbf{a} = (a_i)$. The least squares solution to this system minimizes

$$\|\mathbf{y} - M\mathbf{a}\| = \sqrt{\sum_{i=0}^{d} (y_i - y(x_i))^2}.$$

- Orthonormal bases and the Gram-Schmidt process
- Orthogonal set, orthonormal set, orthogonal basis, orthonormal basis
- An orthogonal set of nonempty vectors in $\mathbb{R}^n$ is linearly independent.

Proof. If $\sum_{i=1}^{k} c_i\mathbf{v}_i = \mathbf{0}$, then $\mathbf{v}_j \cdot \sum_{i=1}^{k} c_i\mathbf{v}_i = c_j\mathbf{v}_j \cdot \mathbf{v}_j = 0$ and $c_j = 0$ for any j.

- If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\mathbf{x} = \sum_{i=1}^{k} (\mathbf{x} \cdot \mathbf{v}_i)\mathbf{v_i}$$

Proof. Let $M = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$. Then, $M^T M = I$.

$$\text{proj}_W\mathbf{x} = M(M^T M)^{-1}M^T\mathbf{x} = MM^T\mathbf{x} = \sum_{i=1}^{k} (\mathbf{v}_i^T\mathbf{x})\mathbf{v_i}.$$

- If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \mathbf{x} = \sum_{i=1}^{k} \frac{\mathbf{x} \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v_i}.$$

- Theorem. Every nonzero subspace of $\mathbb{R}^n$ has an orthogonal basis.

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

  – Let
  $$\mathbf{v}_1 = \mathbf{w}_1.$$

  – Assume that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j\}$ has been constructed with $j < k$. Let $W_j = \text{span}(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j)$. Define
  $$\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \text{proj}_{W_j} \mathbf{w}_{j+1} = \mathbf{w}_{j+1} - \sum_{i=1}^{j} \frac{\mathbf{w}_{j+1} \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \mathbf{v}_i.$$

  – Repeating the above step, we can construct $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, which is an orthogonal basis for W.