Linear Algebra

Chapter 7: Distance and Approximation

University of Seoul
School of Computer Science
Minho Kim

Table of contents

Outline

Outline

Outline

Outline

Best Approximation

- Which vector in a subspace W best approximates (or is closest to) the vector $\boldsymbol{v}$ outside W?

Definition: The Best Approximation

If W is a subspace of a normed linear space V (e.g., $\mathbb{R}^n$) and if $\boldsymbol{v}$ is a vector in V, then the **best approximation to $\boldsymbol{v}$ in W** is the vector $\bar{\boldsymbol{v}}$ in W such that

$$\|\boldsymbol{v} - \bar{\boldsymbol{v}}\| < \|\boldsymbol{v} - \boldsymbol{w}\|$$

for every vector $\boldsymbol{w}$ in W different from $\bar{\boldsymbol{v}}$.

- "shortest distance", "perpendicular distance"

Best Approximation (cont'd)

Theorem 7.8: The Best Approximation Theorem

If W is a finite-dimensional subspace of an inner product space V (e.g., $\mathbb{R}^n$) and if $\boldsymbol{v}$ is a vector in V, then $\mathrm{proj}_W(\boldsymbol{v})$ is the best approximation to $\boldsymbol{v}$ in W.

- What is the **error**? $\|\boldsymbol{v} - \mathrm{proj}_W(\boldsymbol{v})\|$ (distance from $\boldsymbol{v}$ to W)

Least Squares Approximation

- Given data points, (usually obtained from experiments) which function best approximates them? "best fit"
- How to minimize the **error**? (See Figure 7.13 (a) on p.587)

  $\rightarrow$ error vector $\boldsymbol{e} := \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$
- How to define the error?

  "p-norm" $\|\boldsymbol{e}\|_p := \left(\sum_{i=1}^n |\epsilon_i|^p\right)^{\frac{1}{p}}$
  - $\|\boldsymbol{e}\|_1 := \sum_{i=1}^n |\epsilon_i|$
  - $\|\boldsymbol{e}\|_2 := \sqrt{\sum_{i=1}^n |\epsilon_i|^2} = \|\boldsymbol{e}\|$
  - $\|\boldsymbol{e}\|_\infty := \max\left(|\epsilon_1|, \ldots, |\epsilon_n|\right)$
- Least square approximation: Which function best approximates the data points minimizing the **least squares error** $\|e\|$?

Least Squares Approximation (cont'd)

Definition

If A is an $m \times n$ matrix and $\boldsymbol{b}$ is in $\mathbb{R}^m$, a **least squares solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is a vector $\bar{\boldsymbol{x}}$ in $\mathbb{R}^n$ such that

$$\|\boldsymbol{b} - A\bar{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$$

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

Solving Least Squares Problem

- $A\bar{\boldsymbol{x}} \in \mathrm{col}(A)$, therefore, the solution is the closest vector in $\mathrm{col}(A)$ to $\boldsymbol{b}$.
- By the "Best Approximation Theorem", $A\bar{\boldsymbol{x}} = \mathrm{proj}_{\mathrm{col}(A)}(\boldsymbol{b})$.
- $\boldsymbol{b} - A\bar{\boldsymbol{x}} = \boldsymbol{b} - \mathrm{proj}_{\mathrm{col}(A)}(\boldsymbol{b}) = \mathrm{perp}_{\mathrm{col}(A)}(\boldsymbol{b})$ is orthogonal to $\mathrm{col}(A)$
  $\Rightarrow \boldsymbol{b} - A\bar{\boldsymbol{x}}$ is orthogonal to all the columns of A.
  $\Rightarrow A^T(\boldsymbol{b} - A\bar{\boldsymbol{x}}) = \mathbf{0}$
  $\Rightarrow A^T A\bar{\boldsymbol{x}} = A^T\boldsymbol{b}$: A system of **normal equations** for $\bar{\boldsymbol{x}}$

The Least Squares Theorem

Theorem 7.9: The Least Squares Theorem

Let A be an $m \times n$ matrix and let $\boldsymbol{b}$ be in $\mathbb{R}^m$. Then $A\boldsymbol{x} = \boldsymbol{b}$ always has at least one least squares solution $\bar{\boldsymbol{x}}$. Moreover,

a. $\bar{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\bar{\boldsymbol{x}}$ is a solution of the normal equations $A^T A\bar{\boldsymbol{x}} = A^T\boldsymbol{b}$.

b. A has linearly independent columns if and only if $A^T A$ is invertible. In this case, the least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ is unique and is given by

$$\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$$

- Least squares error: $\|\boldsymbol{e}\| = \|\boldsymbol{b} - A\bar{\boldsymbol{x}}\|$

Least Squares via the QR Factorization

When $A \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(A) = n$,

$$\begin{aligned}
A^T A\bar{\boldsymbol{x}} &= A^T\boldsymbol{b} \\
\Rightarrow (QR)^T QR\bar{\boldsymbol{x}} &= (QR)^T\boldsymbol{b} && (A = QR) \\
\Rightarrow R^T Q^T QR\bar{\boldsymbol{x}} &= R^T Q^T\boldsymbol{b} \\
\Rightarrow R^T R\bar{\boldsymbol{x}} &= R^T Q^T\boldsymbol{b} && (Q^T = Q^{-1}) \\
\Rightarrow \bar{\boldsymbol{x}} &= R^{-1}Q^T\boldsymbol{b} && (R \text{ is invertible})
\end{aligned}$$

Theorem 7.10

Let A be an $m\times n$ matrix with linearly independent columns and let $\boldsymbol{b}$ be in $\mathbb{R}^m$. If $A = QR$ is a QR factgorization of A, then the unique least squares solution $\bar{\boldsymbol{x}}$ of $A\boldsymbol{x} = \boldsymbol{b}$ is

$$\bar{\boldsymbol{x}} = R^{-1}Q^T\boldsymbol{b}$$

- Example 7.30 (p.592)
  $\rightarrow$ Don't compute R^{-1}, but solve $R\bar{\boldsymbol{x}} = Q^T\boldsymbol{b}$.

Orthogonal Projection Revisited

Theorem 7.11

Let W be a subspace of $\mathbb{R}^m$ and let A be an $m \times n$ matrix whose columns form a basis for W. If $\boldsymbol{v}$ is any vector in $\mathbb{R}^n$, then the orthogonal projection of $\boldsymbol{v}$ onto W is the vector

$$\mathrm{proj}_W(\boldsymbol{v}) = A(A^T A)^{-1} A^T \boldsymbol{v}$$

The linear transformation $P : \mathbb{R}^m \to \mathbb{R}^n$ that projects $\mathbb{R}^m$ onto W has $A(A^T A)^{-1} A^T$ as its standard matrix.

Proof:

1. By the Least Squares Theorem, the unique least squares solution to $A\boldsymbol{x} = \boldsymbol{v}$ is

$$\bar{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{v}$$

2. And since $A\bar{\boldsymbol{x}} = \mathrm{proj}_{\mathrm{col}(A)}(\boldsymbol{v}) = \mathrm{proj}_W(\boldsymbol{v})$
3. Therefore,

$$\mathrm{proj}_W(\boldsymbol{v}) = A((A^T A)^{-1} A^T \boldsymbol{v}) = (A(A^T A)^{-1} A^T)\boldsymbol{v}$$

Pseudoinverse of a Matrix

- $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ is the unique solution of $A\boldsymbol{x} = \boldsymbol{b}$
- $\bar{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b}$ is the unique least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$

$\rightarrow (A^TA)^{-1}A^T$ plays the role of an "inverse of A"

Definition: Pseudoinverse

If A is a matrix with linearly independent columns, then the **pseudoinverse** of A is the matrix A^+ defined by

$$A^+ = (A^TA)^{-1}A^T$$

- What if A is a square matrix?

Pseudoinverse of a Matrix (cont'd)

- Which properties do they have?

Theorem 7.12

Let A be a matrix with linearly independent columns. Then the pseudoinverse A^+ of A satisfies the following properties, called the **Penrose conditions** for A:

a. $AA^+A = A$

b. $A^+AA^+ = A^+$

c. AA^+ and A^+A are symmetric.

Outline

Outline