Solving the Least-Squares Problem Using Geometry

Objectives

- develop orthogonality condition for the least-squares problem
- find the least-squares problem solution
- introduce matrix inversion

The Least-Squares Problem

N

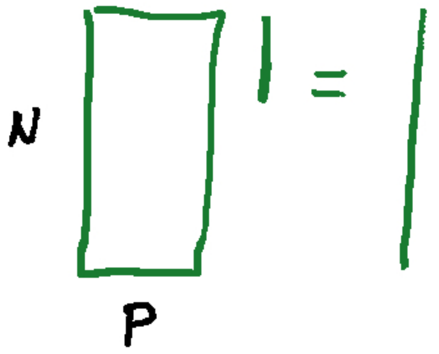
feature vectors

$$\begin{bmatrix} \cdots & \underline{x}_1^T & \cdots \\ \cdots & \underline{x}_2^T & \cdots \\ & \vdots & \\ \cdots & \underline{x}_N^T & \cdots \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

labels

p model parameters

$\underline{A}\,\underline{w} = \underline{d}$



Assume:

- $N \geq P$
- $\text{rank}(\underline{A}) = P$

Goal: $\min_{\underline{w}} \|\underline{A}\,\underline{w} - \underline{d}\|_2^2$

Let $\hat{\underline{d}} = \underline{A}\,\underline{w}$

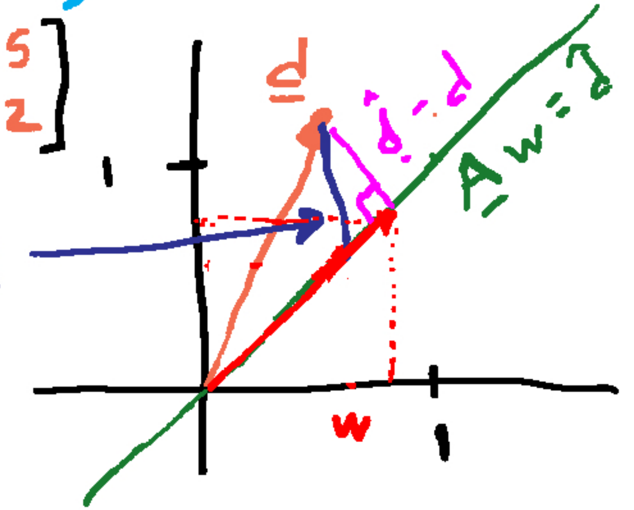
- $\hat{\underline{d}}$ lies in p-dim subspace spanned by columns of $\underline{A}$

Example $(P=1,\ N=2)$

$\underline{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \underline{d} = \begin{bmatrix} 0.5 \\ 1.2 \end{bmatrix}$

$\hat{\underline{d}} = \begin{bmatrix} w \\ w \end{bmatrix}$

$\hat{\underline{d}} - \underline{d}$



Soln: $\hat{\underline{d}} - \underline{d} \perp \underline{A}$

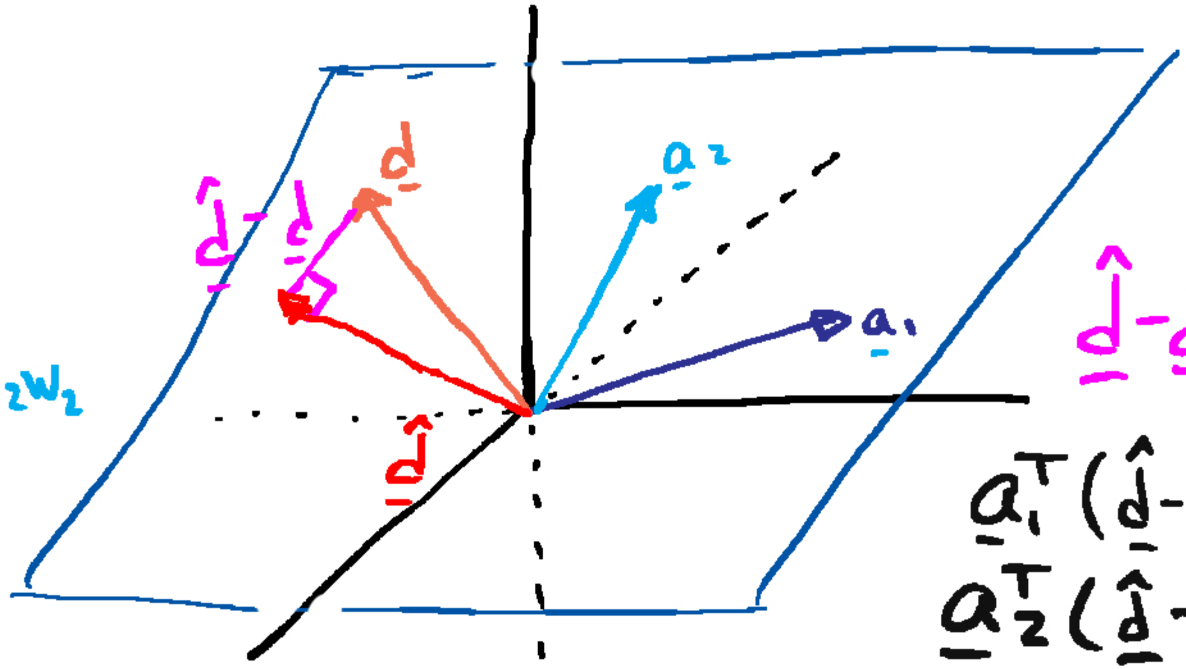
Orthogonality: $\underline{A} \perp (\hat{\underline{d}} - \underline{d})$

$N=3,\ P=2$

$\underline{A} = [\underline{a}_1 \ \ \underline{a}_2]$ $\quad \underline{d}$

$\hat{\underline{d}} = \underline{A}\,\underline{w} = \underline{a}_1 w_1 + \underline{a}_2 w_2$

span$\{\underline{a}_1, \underline{a}_2\}$



$\hat{\underline{d}} - \underline{d} \perp \text{span}\{\underline{a}_1, \underline{a}_2\}$

$\Downarrow$

$\underline{a}_1^T(\hat{\underline{d}} - \underline{d}) = 0$

$\underline{a}_2^T(\hat{\underline{d}} - \underline{d}) = 0$

In general

★ $\underline{A}^T(\hat{\underline{d}} - \underline{d}) = \underline{0}$ "orthognality condition"

Solution: $\underline{A}^T(\underline{A}\,\underline{w} - \underline{d}) = \underline{0} \Rightarrow \underbrace{\underset{N\times P}{\underline{A}^T}\,\underset{P\times N}{\underline{A}}}_{P\times P}\,\underline{w} = \underline{A}^T\underline{d}$

$(\underline{A}^T\underline{A})^{-1}(\underline{A}^T\underline{A})\underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d} \Rightarrow \underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d}$ (matrix inverse)

Matrix Inversion

Let $\underline{B}$ be a $P \times P$ invertible matrix. $\underline{B}^{-1}$ satisfies

$$\underline{B}^{-1}\underline{B} = \underline{I} = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ & & \ddots \\ 0 & & & 1 \end{bmatrix} \quad \text{(identity matrix: } \underline{I}\,\underline{v} = \underline{v}\text{)}$$

$$\underline{B}\,\underline{B}^{-1} = \underline{I}$$

Examples:

$$\underline{B} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \underline{B}^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \Rightarrow \underline{B}\,\underline{B}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\underline{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \underline{B}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \Rightarrow \underline{B}\,\underline{B}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix}$$

Not all matrices have inverses

$$\underline{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \underline{B}^{-1} = \frac{1}{0} \times$$

☆ Full rank (P) square matrices are invertible!

(proof in notes)

A few conditions for invertibility

- $\underline{A}^T\underline{A}$ is invertible iff $\underline{A}$ $(N \times P,\ P \le N)$ is rank P
- Positive definite $\Rightarrow$ invertible. $\underline{Q}$ is positive definite $(\underline{Q} > 0)$ iff $\underline{v}^T\underline{Q}\underline{v} > 0 \ \forall \ \underline{v} \neq 0$ "for all"

  (proofs in notes)

$\underline{A}^T\underline{A}$ is positive definite:

Linearly indepnt $\Rightarrow$ so $y \neq 0$

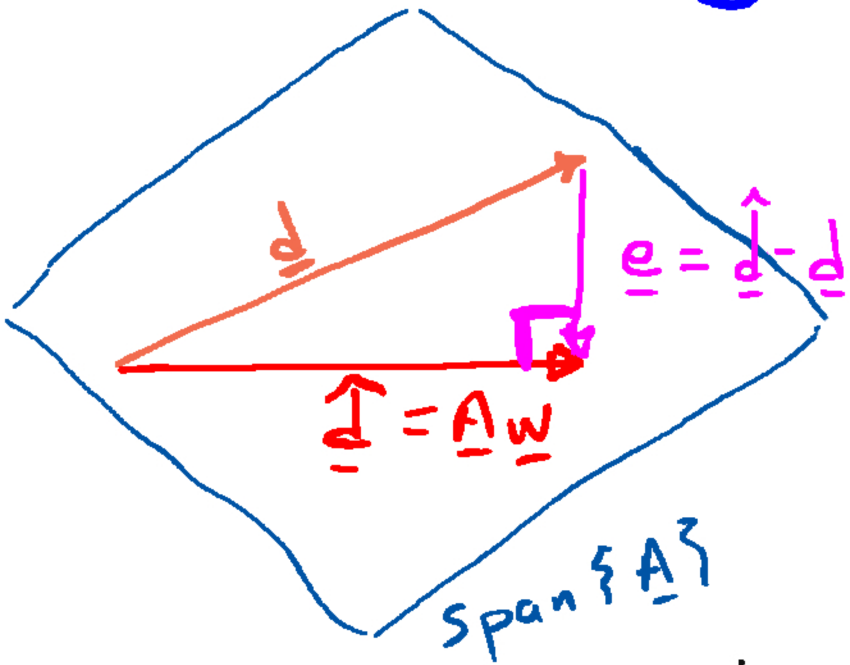
Let $\underline{y} = \underline{A}\underline{v}$. $\text{rank}(\underline{A}) = P \Rightarrow \underline{y} \neq \underline{0}$ for $\underline{v} \neq \underline{0}$

$$(\underline{A}\underline{v})^T\underline{A}\underline{v} = \underline{v}^T\underbrace{\underline{A}^T\underline{A}}_{Q}\underline{v} = \underline{y}^T\underline{y} = \sum_i y_i^2 > 0 \quad (\underline{v} \neq \underline{0})$$

$\Rightarrow (\underline{A}^T\underline{A})^{-1}$ exists

Note: $\underline{Q}$ is positive semidefinite iff $\underline{v}^T\underline{Q}\underline{v} \geq 0 \ \forall \underline{v} \neq \underline{0}$

Summary



$$\min_{\underline{w}} \|\underline{A}\,\underline{w}-\underline{d}\|_2^2 \Rightarrow \min_{\underline{w}} \|\underline{e}\|_2^2$$

$$\Rightarrow \underline{e} \perp \text{span}\{\underline{A}\} \qquad \underline{A}^T\underline{e} = \underline{0}$$

$$\underbrace{\underline{A}^T}_{P \times N}\Big(\underbrace{\underline{A}}_{N \times P}\,\underbrace{\underline{w}}_{P \times 1} - \underbrace{\underline{d}}_{N}\Big) = \underline{0}$$

$$\underbrace{\underline{A}^T\underline{A}}_{P \times P}\,\underbrace{\underline{w}}_{P\times 1} = \underbrace{\underline{A}^T\underline{d}}_{P\times 1}$$

$$\boxed{\underline{w} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d}}$$

$$\underline{A}\,\underline{w} \rightarrow (\underline{A}^T\underline{A})^{-1}\underline{A}^T\,\underline{A}\,\underline{w} = \underline{w}$$

$$\hat{\underline{d}} = \underline{A}\overbrace{(\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d}}^{w}$$

Solving the Least-Squares Problem Using Geometry Proofs: Invertibility of Full Rank and Positive Definite Matrices

Background: A matrix is rank P if there are P linearly independent columns (or rows). Vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_P$ are linearly independent if and only if

$$\sum_{i=1}^{P} \boldsymbol{v}_i c_i = \mathbf{0}$$

implies $c_i = 0, i = 1, 2, \ldots, c_P$. For convenience we may write this condition in matrix vector form as $\boldsymbol{V}\boldsymbol{c} = \mathbf{0}$ if and only if $\boldsymbol{c} = \mathbf{0}$ where

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_P \end{bmatrix}$$

and $\boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & \ldots & c_P \end{bmatrix}^T$.

The columns of a rank P, N-by-P matrix span a P-dimensional subspace of $\mathbb{R}^N$.

1. **$\boldsymbol{A}^T\boldsymbol{A}$ is full rank.** If $\boldsymbol{A}$ is N-by-P with $P \leq N$ and rank$\{\boldsymbol{A}\} = P$, then $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A}$ (P-by-P) has rank P.

   We will prove this result by contradiction. Suppose $\boldsymbol{B}$ has rank less than P. This implies there is a vector $\boldsymbol{v} \neq \mathbf{0}$ such that $\boldsymbol{B}\boldsymbol{v} = \mathbf{0}$. Since $\boldsymbol{B}\boldsymbol{v} = \mathbf{0}$, we have $\boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} = 0$ for some $\boldsymbol{v} \neq \mathbf{0}$. Note that $\boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v} = (\boldsymbol{A}\boldsymbol{v})^T\boldsymbol{A}\boldsymbol{v} = \boldsymbol{y}^T\boldsymbol{y}$ where we define $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{v}$. But rank$\{\boldsymbol{A}\} = P$, so there is no $\boldsymbol{v} \neq \mathbf{0}$ for which $\boldsymbol{A}\boldsymbol{v} = \mathbf{0}$. Thus, if $\boldsymbol{v} \neq \mathbf{0}$, then $\boldsymbol{y} \neq \mathbf{0}$. If $\boldsymbol{y} \neq \mathbf{0}$, then $\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} > 0$. Hence there is a contradiction and rank$\{\boldsymbol{B}\} = P$.

2. **Positive Definite Matrices are Full Rank.** If a P-by$P$ matrix $\boldsymbol{Q}$ is positive definite, that is, $\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} > 0$ for all $\boldsymbol{v} \neq \mathbf{0}$, then rank$\{\boldsymbol{Q}\} = P$.

   The proof is similar to the previous one. Suppose rank$\{\boldsymbol{Q}\} < P$. This implies there is a vector $\boldsymbol{v} \neq \mathbf{0}$ such that $\boldsymbol{Q}\boldsymbol{v} = \mathbf{0}$. If $\boldsymbol{Q}\boldsymbol{v} = \mathbf{0}$, then $\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} = \boldsymbol{v}^T\mathbf{0} = 0$, which contradicts the assumption that $\boldsymbol{Q}$ is positive definite. Thus rank$\{\boldsymbol{Q}\} = P$.

3. **Full Rank Square Matrices Are Invertible.** If $\boldsymbol{B}$ is P-by-P and rank$\{\boldsymbol{B}\} = P$, then $\boldsymbol{B}$ is invertible. That is, there exists $\boldsymbol{B}^{-1}$ satisfying $\boldsymbol{B}\boldsymbol{B}^{-1} = \boldsymbol{I}$.

   Since rank$\{\boldsymbol{B}\} = P$, the columns of $\boldsymbol{B}$ span $\mathbb{R}^P$ and any vector $\boldsymbol{z} \in \mathbb{R}^P$ can be written as a weighted combination of the columns of $\boldsymbol{B}$, that is, $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{v}$.

   Now, let $\boldsymbol{e}_i \in \mathbb{R}^P$ be the vector of all zeros except for a one in the i^{th} row, e.g., $\boldsymbol{e}_2 = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$. Since the columns of $\boldsymbol{B}$ span $\mathbb{R}^P$, there is a vector $\boldsymbol{v}_i$ so that $\boldsymbol{e}_i = \boldsymbol{B}\boldsymbol{v}_i, i = 1, 2, \ldots, P$. Concatenating these relationships for $i = 1, 2, \ldots, P$ gives

$$\left[\begin{array}{cccc} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_P \end{array}\right] = \boldsymbol{B}\left[\begin{array}{cccc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_P \end{array}\right]$$

But $\left[\begin{array}{cccc} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_P \end{array}\right] = \boldsymbol{I}$, so thus by definition $\left[\begin{array}{cccc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_P \end{array}\right] = \boldsymbol{B}^{-1}$.