EE445 Mod2-Lec1: Introduction to Least Squares

References:

- [VMLS]: Chapter 12

Least Squares Set-up

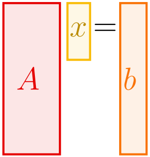
- Linear Regression ([VMLS, Ch. 2.3]) is the simplest form of machine learning out there.
- Consider an $m \times n$ matrix A—i.e., $A \in \mathbb{R}^{m\times n}$—and vectors $b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$
- **Notation**:
- **Goal**:

Least Squares Set-up: ML Interpretation

- **Goal**: Find a solution to $Ax = b$—that is, find x such that $Ax = b$

Overdetermined System of Equations→Least Squares Opt

- **Goal**: Find a solution to $Ax = b$—that is, find x such that $Ax = b$



Least Squares Optimization Problem

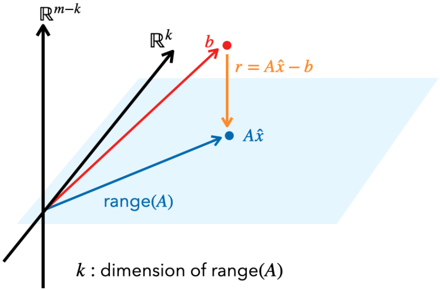
- Least squares optimization problem:

(LS prob)

Least Squares Optimization Problem

- Least squares optimization problem:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$



- **residual**:

Example Applications: Advertising Purchases

Other Examples

- Stock market prediction:
- Weather forecasting:
- Predicting impact of GPA/SAT scores on college admissions
- Predicting/forecasting housing prices as a function of size, location, etc.

Combing back to the optimization problem

- Any vector $\hat{x}$ satisfying the following is a solution (i.e., a least squares approximate solution):

Column Interpretation

$$A = \begin{bmatrix} | & \cdots & | \\ a_1 & \cdots & a_n \\ | & \cdots & | \end{bmatrix}, \quad a_i \in \mathbb{R}^m$$

Row Interpretation

$$A = \begin{bmatrix} - & \tilde{a}_1^\top & - \\ \vdots & \dots & \vdots \\ - & \tilde{a}_m^\top & - \end{bmatrix}, \quad \tilde{a}_i \in \mathbb{R}^n$$

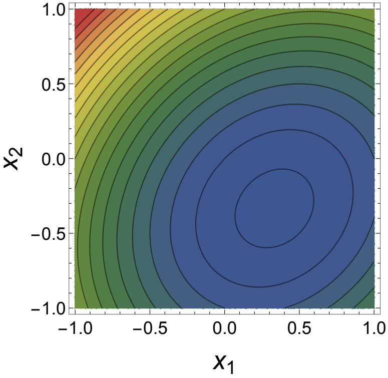
Example

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad m = 3,\ n = 2$$

Aside: Finding Minima via Calculus

Example Continued

$$\min_{x_1,x_2}\{(2x_1-1)^2+(-x_1+x_2)^2+(2x_2+1)^2\}$$



Least Squares Solution via Calculus

Assumption [A1]: The columns of A are linearly independent—i.e., $\sum_{i=1}^{n} c_i a_i = 0 \iff c_i = 0 \quad \forall i = 1, \ldots, n$

Let's verify

- Least squares objective in summation form:

$$f(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2$$

Least Squares Solution via Calculus Continued

- Any minimizer $\hat{x}$ of $f(x) = \|Ax - b\|_2^2$ must satisfy

- **Gram matrix**: $A^\top A$ has entries which are the inner products of the columns of A
- **[A1]** $\implies$ $A^\top A$ is invertible [VMLS, §11.5,pg. 214]
- Hence, $\hat{x} = (A^\top A)^{-1} A^\top b$ is the *only* solution of the normal equations
- **Pseudo-inverse**: $A^\dagger := (A^\top A)^{-1} A^\top$ is a left inverse of A
- $\hat{x} = A^\dagger b$ solves $Ax = b$ **if** the set of equations has a solution otherwise it is said to be the least squares approximate solution.

Direct Verification of the Solution

- Let's check via direct verification: we will show that for any $x \neq \hat{x} = A^\dagger b$ we have the estimate

$$\|A\hat{x} - b\|_2^2 < \|Ax - b\|_2^2$$

- Indeed,

$$\begin{aligned} \|Ax - b\|_2^2 &= \|(Ax - A\hat{x}) + (A\hat{x} - b)^2\|_2^2 \\ &= \|A(x - \hat{x})\|_2^2 + \|A\hat{x} - b\|_2^2 + 2(x - \hat{x})^\top A^\top (A\hat{x} - b) \end{aligned}$$

  since $\|u + v\|_2^2 = (u + v)^\top (u + v) = \|u\|_2^2 + \|v\|_2^2 + 2u^\top v$

- **Claim**: $(x - \hat{x})^\top A^\top (A\hat{x} - b) = 0$
  proof: since $(A^\top A)\hat{x} = A^\top b$ [normal equations], we have

$$(x - \hat{x})^\top A^\top (A\hat{x} - b) = (x - \hat{x})^\top (A^\top A\hat{x} - A^\top b) = 0$$

Direct Verification of the Solution

- we know that $(x-\hat{x})^\top A^\top (A\hat{x}-b)=0$
- Coming back to the expression for the objective, we have

$$\|Ax-b\|_2^2 = \underbrace{\|A(x-\hat{x})\|_2^2}_{\geq 0} + \|A\hat{x}-b\|_2^2$$

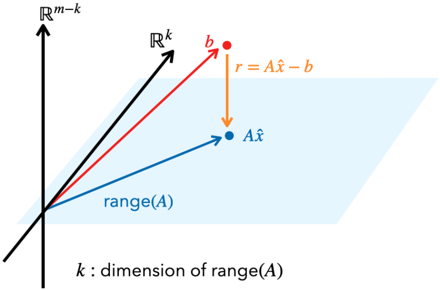
- Hence, we deduce

$$\|A\hat{x}-b\|_2^2 \leq \|Ax-b\|_2^2$$

- **Row form of solution**: sometimes its useful to express the solution as

$$\hat{x} = A^\dagger b = (A^\top A)^{-1}A^\top b = \left(\sum_{i=1}^m \tilde{a}_i\tilde{a}_i^\top\right)^{-1}\left(\sum_{i=1}^m b_i\tilde{a}_i\right)$$

Orthogonality Principle

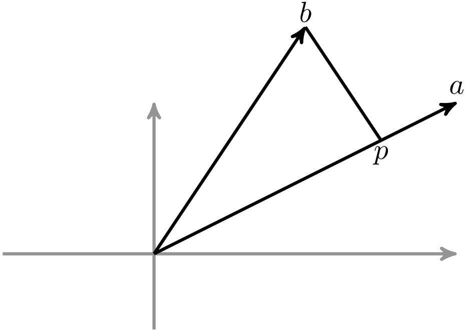


- $A\hat{x}$ is the linear combination of columns of A closest to b
- Residual $r = A\hat{x} - b$ satisfies the so orthogonality principle:

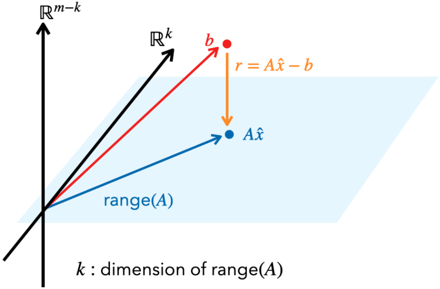
$$(Az) \perp r \quad \forall\ z \in \mathbb{R}^n$$

- **Why?**

Let's Look at the Vector case



Orthogonality Principle



- $A\hat{x}$ is the linear combination of columns of A closest to b
- Residual $r = A\hat{x} - b$ satisfies the so orthogonality principle:

$$(Az) \perp r \quad \forall\, z \in \mathbb{R}^n$$

- **Why?**

- First, [normal equations] $\iff A^\top(A\hat{x} - b) = 0$
- Hence, for any $z \in \mathbb{R}^n$, we have

$$(Az)^\top r = (Az)^\top(A\hat{x} - b) = z^\top A^\top(A\hat{x} - b) = 0$$

Projection

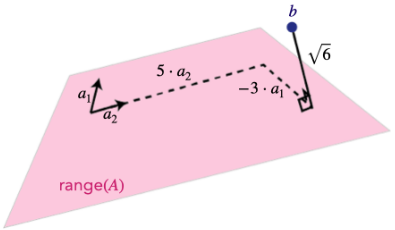
More Examples

Consider

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

Find the least squares approximate solution to $Ax = b$.
Solution.

Example Continued



Numerical Methods of Finding Solutions

Numerical Examples