Linear Regression

EE445 Mod2-Lec1: Introduction to Least Squares

References:

- [VMLS]: Chapter 12

Least Squares Set-up

$\mathbb{R}$ is set of real numbers

$\mathbb{R}^{m\times n}$

$\mathbb{R}^m = \mathbb{R} \times \cdots \times \mathbb{R}$

- 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**: Find a solution to $Ax = b$

ML Interpretation:

- A is a matrix of training data where m : # of samples & n : # of "features"
- $b \in \mathbb{R}^m$: m target values
- $x \in \mathbb{R}^n$: feature weights

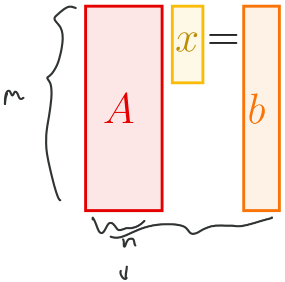
Least Squares Set-up: ML Interpretation

- **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 $\underline{Ax = b}$

$m >> n$



- However, typically A is "tall" → "overdetermined system"
- There is often not an exact soln → we define an "optimization problem" to find an "approximate" soln.

"Least squares Approx. soln."

Least Squares Optimization Problem

$x = (x_1, \ldots, x_n)$

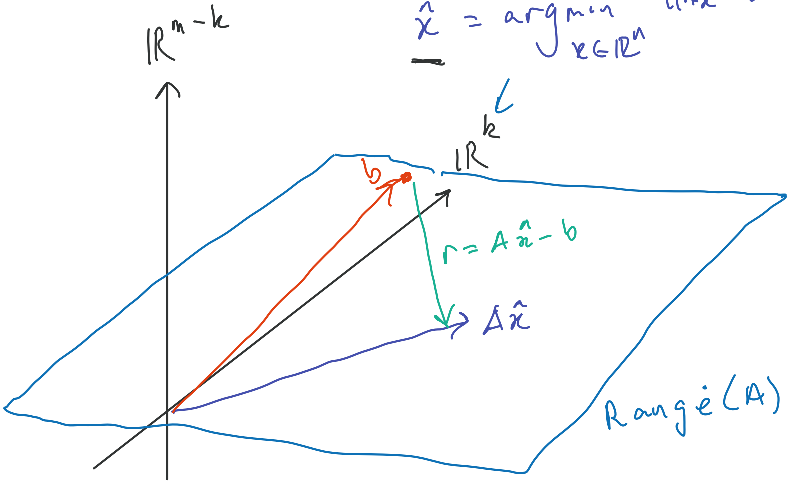
$\|x\|_1 = |x_1| + \cdots + |x_n|$

$\|x\|_2^2 = \left(\sqrt{x_1^2 + \cdots + x_n^2}\right)^{1/2}$

- Least squares optimization problem:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 \qquad \text{(LS prob)}$$

$$\hat{x} = \operatorname{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|^2$$



k: dimension of range(A)

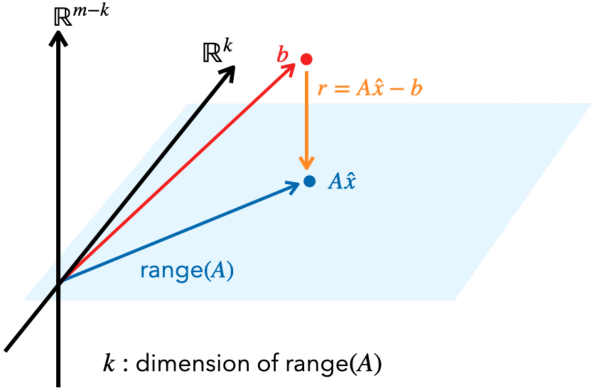
- Components of optimization prob.
  - decision variable: $x \in \mathbb{R}^n$
  - data: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$
  - objective: $\|Ax - b\|^2$
- vector of Residuals $r \in \mathbb{R}^m$

$r = A\hat{x} - b$

Least Squares Optimization Problem

- Least squares optimization problem:

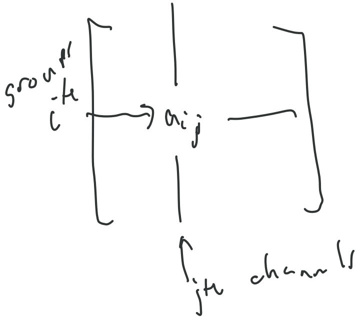
$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$



- **residual**:

Example Applications: Advertising Purchases

- m demographic groups (audiences) that we want to advertise to
- with a target # of 'impressions' or views : b ←
- n different "channels" : $x \in \mathbb{R}^n$ is set of weights "spending"
- $A \in \mathbb{R}^{m\times n}$: a_{ij} = # impressions in group i per \$ spent on channel j



<u>goal:</u> find x such that $\|Ax - b\|^2$

spending

target views

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):

$$\hat{x} \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - b\|^2$$

$$\Big[\underbrace{\|A\hat{x} - b\|^2}_{\|r\|^2} \leq \|A\underline{x} - b\|^2 \quad \text{for any other } x \in \mathbb{R}^n \Big]$$

- It may <u>not</u> be the case that $A\hat{x} = b$
- Regression: $\hat{x}$ is the result of regressing b onto columns of A

Column Interpretation

$$A = \left.m\right\{\underbrace{\begin{bmatrix} | & \cdots & | \\ a_1 & \cdots & a_n \\ | & \cdots & | \end{bmatrix}}_{n}, \quad a_i \in \mathbb{R}^m$$

$x_i \cdot a_i$

$$2 \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

- Lst Sqrs Prob $\equiv$ finding linear combo. of cols. that is closest to b

$$\|Ax - b\|^2 = \|x_1 \cdot a_1 + x_2 \cdot a_2 + \cdots + x_n \cdot a_n - b\|^2$$

- For a soln. $\hat{x}$, $\quad A\hat{x} = \sum_{i=1}^{n} \hat{x}_i \cdot a_i$

- $A\hat{x}$ is the "closest" to b among all linear combos of cols. of A

Row Interpretation

$$\left[- \; \tilde{a}_i^\top \; - \right]^\top = \begin{bmatrix} | \\ \tilde{a}_i \\ | \end{bmatrix} \in \mathbb{R}^n$$

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

- $r = A\hat{x} - b$
- components of r: $r_i = \tilde{a}_i^\top \hat{x} - b_i, \quad i = 1, \dots, m$
- objective: $\|Ax - b\|^2 = (\tilde{a}_1^\top x - b_1)^2 + \dots + (\tilde{a}_m^\top x - b_m)^2$

Example

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad m = 3,\ n = 2$$

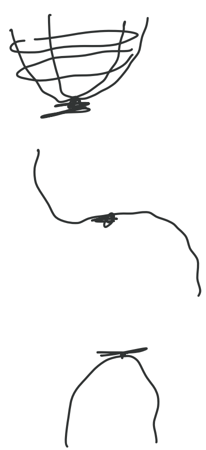
- $Ax = b \Rightarrow \{ 2x_1 = 1,\ -x_1 + x_2 = 0,\ 2x_2 = -1 \} \rightarrow$ no soln.

- row version:

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

[VMLS, App. C]

Aside: Finding Minima via Calculus



- CALCULUS: Consider $f(x)$. to find $\min_x f(x)$

scalar:

- First order condition: $\frac{d}{dx} f(x) = 0 \Rightarrow x^*$ soln.
- 2^{nd} order cond.: $\left.\frac{d^2}{dx^2} f(x)\right|_{x=x^*} > 0$

Multivariable: $\nabla f(x) = 0$ (1^{st} order) $\to x^*$ soln.

$$\nabla f(x) = \begin{bmatrix} \nabla_1 f(x) \\ \vdots \\ \nabla_n f(x) \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} > 0 \iff \nabla^2 f(x^*) \text{ has positive eigenvalues}$$

$$\nabla_i f(x) = \frac{\partial}{\partial x_i} f(x)$$

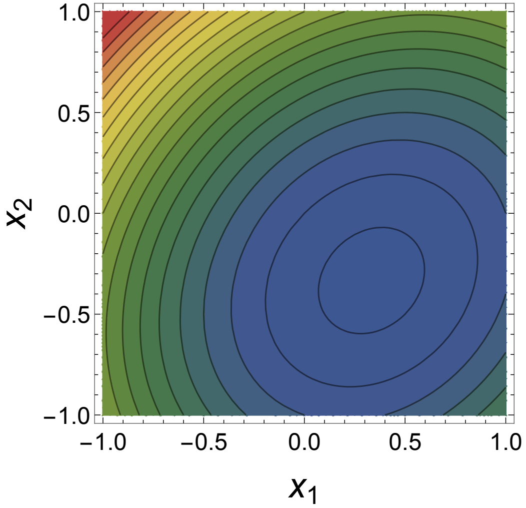
Example Continued

$f: \mathbb{R}^2 \to \mathbb{R}$

$\nabla f(x) = 0$

$\|Ax-b\|^2$

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



$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2(2x_1-1) - 2(-x_1+x_2) \\ 2(-x_1+x_2) + 2(2x_2+1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \hat{x} = \begin{bmatrix} 1/3 \\ -1/3 \end{bmatrix} \leftarrow$$

- $A\hat{x} \neq b$
- $r = A\hat{x} - b = \left(-\frac{1}{3}, -\frac{2}{3}, \frac{1}{3}\right)$

$\|A\hat{x}-b\|^2 = \frac{2}{3}$

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

- A minimizer $\hat{x}$ of $\underline{f(x)} = \|Ax - b\|^2$ | $\nabla f(x) = 2A^T(Ax - b)$

$$\nabla f(\hat{x}) = \underbrace{2A^T(A\hat{x} - b)} = 0 \iff A^TA\hat{x} = A^Tb \quad \text{[normal eqns.]}$$

- Gram matrix : $\underline{A^TA}$

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}$$

- [A1] colms. of A are linearly independent $\Rightarrow$ A^TA is invertible

$$\Rightarrow \quad \underbrace{(A^TA)^{-1}A^TA}\hat{x} = (A^TA)^{-1}A^Tb \Rightarrow \hat{x} = \underbrace{(A^TA)^{-1}A^T}_{\overset{!!}{A^\dagger}} b$$

pseudo-inverse

Direct Verification of the Solution

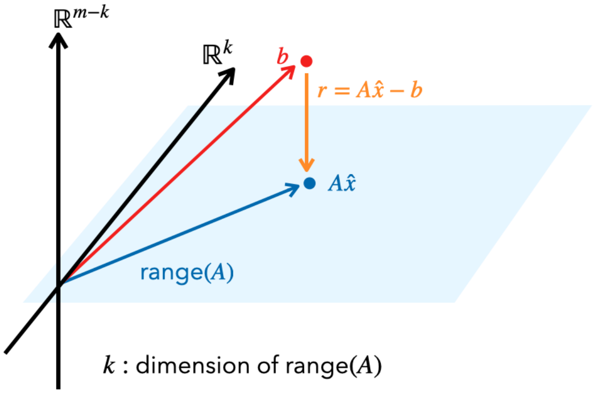
- Let's check via direct verification: we will show that for any $x \neq \hat{x} = A^\dagger b$ we have the estimate
$$\|A\hat{x} - b\|_2^2 < \|Ax - b\|_2^2$$
- Indeed,

Direct Verification of the Solution

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

Orthogonality Principle

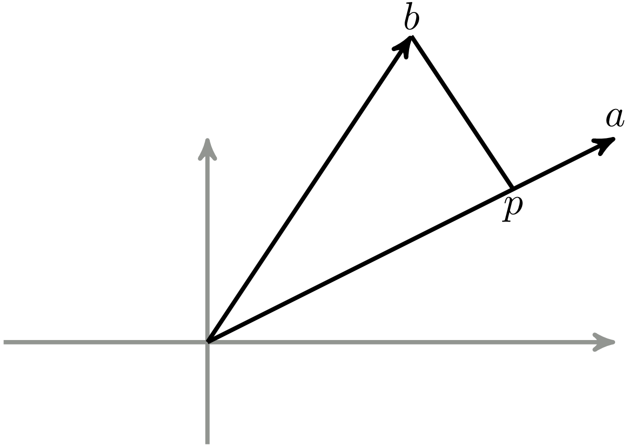


- $A\hat{x}$ is the linear combination of columns of A closest to b
- Residual $r = A\hat{x} - b$ satisfies the so orthogonality principle:

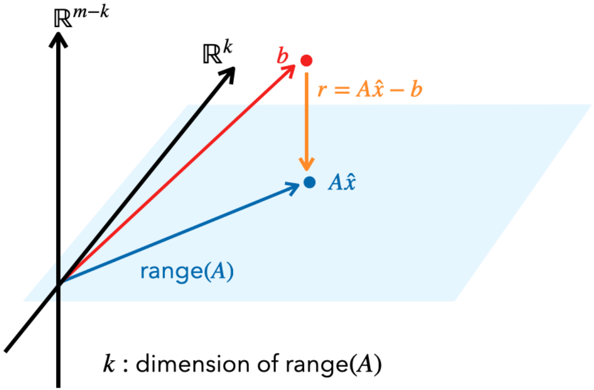
$$(Az) \perp r \quad \forall\ z \in \mathbb{R}^n$$

- **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

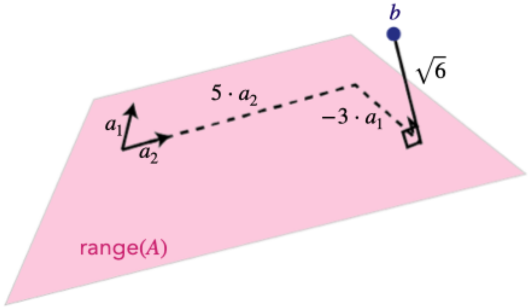
Consider

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

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

Solution.

Example Continued



Numerical Methods of Finding Solutions

Numerical Examples