Least Squares: Part 2

Lecture 3-2 - CMSE 382

Prof. Elizabeth Munch

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Fri, Jan 30, 2026

Topics covered

- Regularized least squares
- Tikhonov regularization
- De-noising

Section 1

Regularized Least Squares

Ordinary Least Squares Approximation

Definition

The **residual sum of squares error** (or total sum of square errors) is

$$\text{RSS} = \|A\mathbf{x}_{\text{LS}} - \mathbf{b}\|^2$$

The minimizer of this equation is the **least squares estimate**

$$\mathbf{x}_{\text{LS}} = (\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top \mathbf{b}$$

Regularized least squares

Definition

We add a **regularization function** $R(\cdot)$ to OLS to obtain the **regularized least squares** function:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|^2 + \lambda R(\mathbf{x})$$

- $\lambda > 0$ determines the weight given to the regularization function.
- R is chosen based on prior knowledge or desired behavior
- When $R(\mathbf{x}) = 0$, we recover ordinary least squares.
- When $R(\mathbf{x}) = \|\mathbf{x}\|_2^2 = \sum_{i=1}^n x_i^2$, we have ridge regression.
- When $R(\mathbf{x}) = \|\mathbf{x}\|_1$, we have LASSO regression.

Regularized least squares

Why add R?

- Can solve ill-posed problems.
  - Underdetermined systems have infinite solutions and OLS fails. Regularized least squares allows constraining the problem.
- OLS focuses on reducing the error sum of squares, but can lead to poor fits in-between data points.
  - A regularization term can penalize big errors in-between data points for a better fit.
- Regularization allows including prior knowledge into the model:
  - E.g., allows smoother fits for noisy data (denoising)

Tikhonov (Quadratic) Regularization

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|^2 + \lambda R(\mathbf{x}) \qquad \Rightarrow \qquad \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|^2 + \lambda\|D\mathbf{x}\|^2$$

Definition

When $R(\mathbf{x}) = \|D\mathbf{x}\|^2$, we have **Tikhonov regularization**. D is called the **Tikhonov matrix**.

- We require that the null space of D must intersect the null space of A at $\mathbf{0}$ for a unique solution.
- Often D is a scalar multiple of the identity matrix.

Solving Tikhonov Regularization

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|^2 + \lambda\|D\mathbf{x}\|^2 = \mathbf{x}^\top(A^\top A + \lambda D^\top D)\mathbf{x} - 2\mathbf{b}^\top A\mathbf{x} + \|\mathbf{b}\|^2$$

If $\nabla^2 f = 2(A^\top A + \lambda D^\top D) \succ 0$, then $x_{\mathrm{LS}} = (A^\top A + \lambda D^\top D)^{-1}A^\top\mathbf{b}$

Tikhonov Regularization Choices

Order	$R(\mathbf{x})$	Promotes
zeroth	$\lVert L_0\,\mathbf{x}\rVert^2$	small $\lVert\mathbf{x}\rVert$, L_0 is the identity matrix
First	$\lVert L_1\,\mathbf{x}\rVert^2$	smoothness by minimizing 1st derivative
Second	$\lVert L_2\,\mathbf{x}\rVert^2$	smoothness by minimizing 2nd derivative

$$x'|_i \approx \frac{x_{i+1} - x_i}{h}$$

$$L_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

$$x''|_i \approx \frac{x_{i+1} - 2x_i + x_{i-1}}{h^2}$$

$$L_2 = \begin{bmatrix} -2 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 \end{bmatrix}$$

Noise

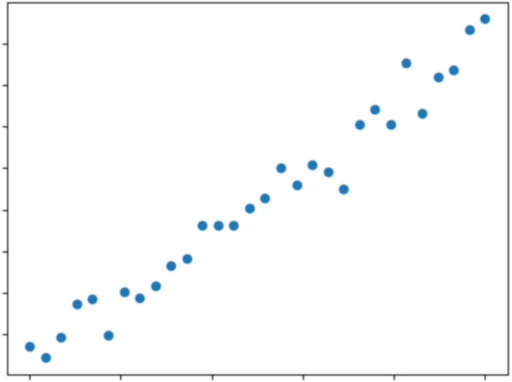
Noisy data can be written as

$$\mathbf{b} = \mathbf{x} + \mathbf{w}$$

where $\mathbf{x}$ is the signal vector and $\mathbf{w}$ is the noise vector.

Problem Statement

Given $\mathbf{b}$, find a good estimate for $\mathbf{x}$



Denoising using Regularized Least Squares

Idea: Directly incorporate prior knowledge that the solution is smooth.

- Small difference between any two subsequent function values
- Popular choice:
  $$R(\mathbf{x}) = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2$$

- Can write as a Tikhonov matrix:
  $$R(\mathbf{x}) = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 = \|L\,\mathbf{x}\|^2$$

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

- The regularized least squares problem becomes $\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|^2$, with solution $\mathbf{x}_{\text{RLS}}(\lambda) = (\mathbf{I} + \lambda \mathbf{L}^\top \mathbf{L})^{-1}\mathbf{b}$

Groups - Round 2

Group 1
Abigail, Shanze, Jack, Quang Minh,

Group 2
Igor, Atticus, K M Tausif, Long,

Group 3
Yousif, Zheng, Jake, Purvi,

Group 4
Maye, Alice, Arjun, Kyle,

Group 5
Monirul Amin, Jay, Brandon, Luis,

Group 6
Scott, Ha, Lora, Tianjian,

Group 7
Braedon, Sai, Joseph, Noah,

Group 8
Michal, Aidan, Jonid, Dev,

Group 9
Vinod, Saitej, Anthony, Breena,

Group 10
Karen, Dori, Lowell, Aaron,

Group 11
Jamie, Sanskaar, Dominic, Lauryn,

Group 12
Andrew, Arya, Daniel, Morgan,

Next time

- Check course webpage for videos and reading for next class
- Office hours posted on the course webpage
- Homework posted, due **TODAY**, Friday, Jan 30 at 11:59pm
- Quiz 2 on Weds 2/4