1 Least Squares

PVK 2019: MAD

bacdavid@student.ethz.ch

gitlab.ethz.ch/bacdavid

Schedule

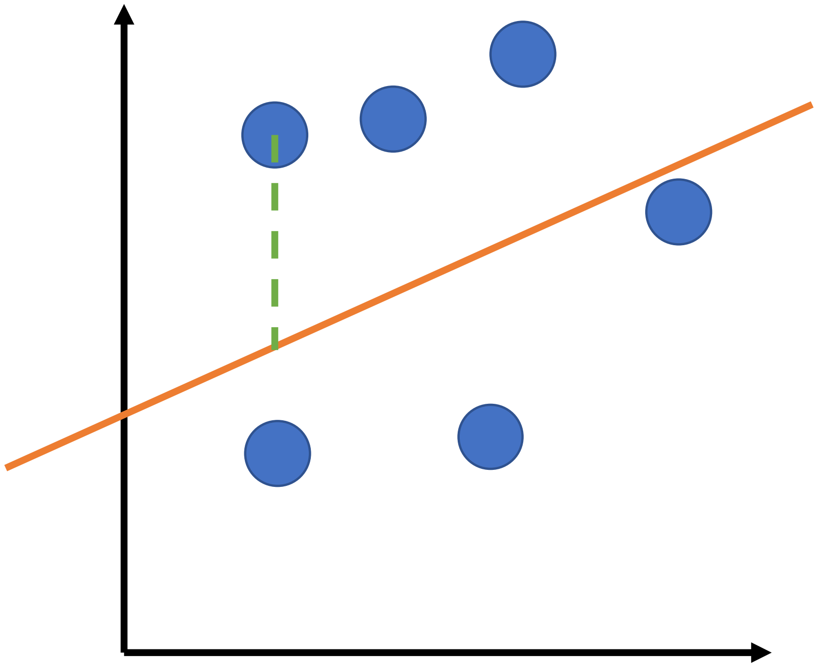
1. Theory
   1. Setup
   2. LLS in 1D
   3. LLS in KD
   4. SVD
   5. Pseudo inverse
   6. LLS with large condition number
2. Exercises
3. Homework

Theory

Setup

- Dataset in 1D $D_1 = \{(x_1, y_1), \dots, (x_N, y_N)\}$
- Dataset in nD $D_n = \{(\overrightarrow{x_1}, y_1), \dots, (\overrightarrow{x_N}, y_N)\}$
- Datavector $\vec{x} = [x_1, \dots, x_N]^T$
- Feature $\overrightarrow{x_i} = \left[x_1^i, \dots, x_n^i\right]^T$
- Datamatrix $X = [\overrightarrow{x_1}, \dots, \overrightarrow{x_N}]^T$
- Labelvector $\vec{y} = [y_1, \dots, y_N]^T$
- Weight/ Parameter w
- Weightvector $\vec{w} = [w_1, \dots, w_K]^2$
- 2-Norm $\|\cdot\|_2$

LLS in 1D



- Datapoint
- LLS fit „prediction“
- Error

LLS in 1D cont.

- Problem formulation:

$$\vec{x}w \approx \vec{y}$$

- Error between LHS and RHS:

$$e = \vec{x}w - \vec{y}$$

- Minimize the squared error to obtain optimal weight:

$$w^* = \arg\min_{w} e^2 = \arg\min_{w} ||\vec{x}w - \vec{y}||_2^2$$

LLS in 1D cont. Example

- Compute the closed form solution to $w^* = \arg\min_{w} \|\vec{x}w - \vec{y}\|_2^2$.

LLS in 1D cont. Example

- Given:
  - $D_1 = \{(0,0),(1,1.5),(2,2.2)\}$

- Task:
  - Set up all vectors
  - Compute the LLS solution manually; Solve $\arg\min_w \|\vec{x}w - \vec{y}\|_2^2$ (don‘t plug in)
  - Predict a the label for $x = 10$

LLS in KD

- Problem formulation:

$$X\vec{w} = \vec{y}$$

- Error between LHS and RHS:

$$e = X\vec{w} - \vec{y}$$

- Minimize the squared error to obtain optimal weight:

$$\vec{w}^* = \arg\min_{\vec{w}} e^2 = \arg\min_{\vec{w}} \|X\vec{w} - \vec{y}\|_2^2$$

LLS in KD cont.

- Closed form solution to $\vec{w}^* \arg\min_{\vec{w}} \|X\vec{w} - \vec{y}\|_2^2$:

$$\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$$

- Also called “normal equation”

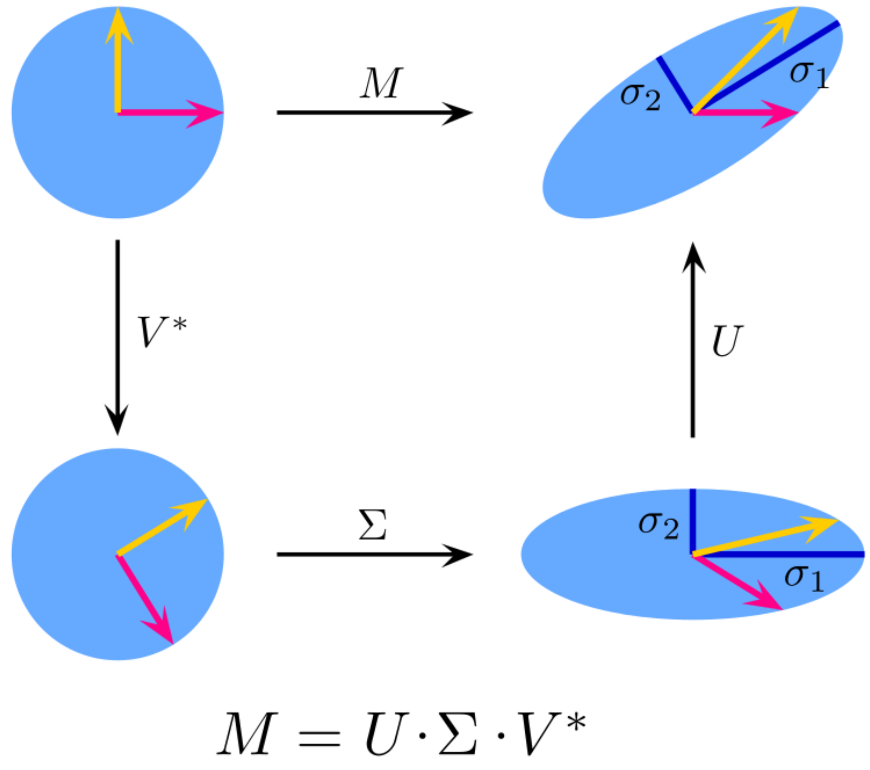
LLS in KD cont. Example

- Given:
  - $D_1 = \{(1,\ 1), (2,\ 3), (3,\ 5)\}$
  - Proposed model: $w_1 + w_2 \cdot x + w_3 \cdot x^2 = y$

- Task:
  - Set up all vectors and matrices, $X\vec{w} = \vec{y}$
  - Compute the solution $\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$ (use a computer)
  - Predict a the label for $\vec{x} = 5$

SVD

- U, V: Orthogonal matrices
- Σ: non-squared diagonal Matrix
- Decomposition:
$$M = U\Sigma V^T$$
- Inverse:
$$\mathrm{M}^{-1} = \mathrm{V}\Sigma^{-1}\mathrm{U}^{\mathrm{T}}$$
- Note: $diag^{-1}(\sigma_i\ ) = diag\left(\frac{1}{\sigma_i}\right)$



Pseudo inverse

- Remember SVD:

$$M^{-1} = V\Sigma^{-1}U^T$$

- If $\sigma_i = 0 \iff M$ is singular
- If M is non-square the inv. doesn't exist
- Define the pseudo inverse:
  - Transpose Σ
  - Reciprocal of non-zero singular values
  - Zeros remain in Σ
  - $M^+ = V\Sigma^+U^T$

$$\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

LLS with large condition number

- Condition number of a system $A\vec{x} = \vec{b}$:

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

- Describes the impact of numerical error:

$$\|\vec{x}_{true} - \vec{x}_{num}\| \sim \kappa \cdot \left\|\vec{b}_{true} - A\vec{x}_{num}\right\|$$

- Problem is “well” conditioned if κ not too large: **Use normal equation**
- Problem is “ill” conditioned if is large: **Use pseudo inverse**

Exercises

Exercise 1

- Given:
  - $D_2 = \{([1,1]^T,\ 1), ([1,2]^T,\ 3), ([2,4]^T,\ 5)\}$
  - Proposed model: $w_1 + w_2 \cdot x_1 + w_3 \cdot x_2^2 = y$

- Task:
  - Set up all vectors and matrices, $X\vec{w} = \vec{y}$ (how to incorporate w_1 and x_2^2 ?)
  - Compute the solution $\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$ (use a computer)
  - Predict a the label for $\vec{x} = [10, 10]^T$

Exercise 2

- Given:
  - Dataset D_1
  - Proposed model: $w_1 \exp(w_2 x + w_3 x^2) = y$

- Task:
  - Set up the LLS formulation

Exercise 3

- Given:
  - $D_{3,x} = \{[1,\ 0,\ 0], [0,\ 2,\ 0], [0, 0, ½]\}$
- Task:
  - Compute the conditional number
  - Is the problem well conditioned?

Homework

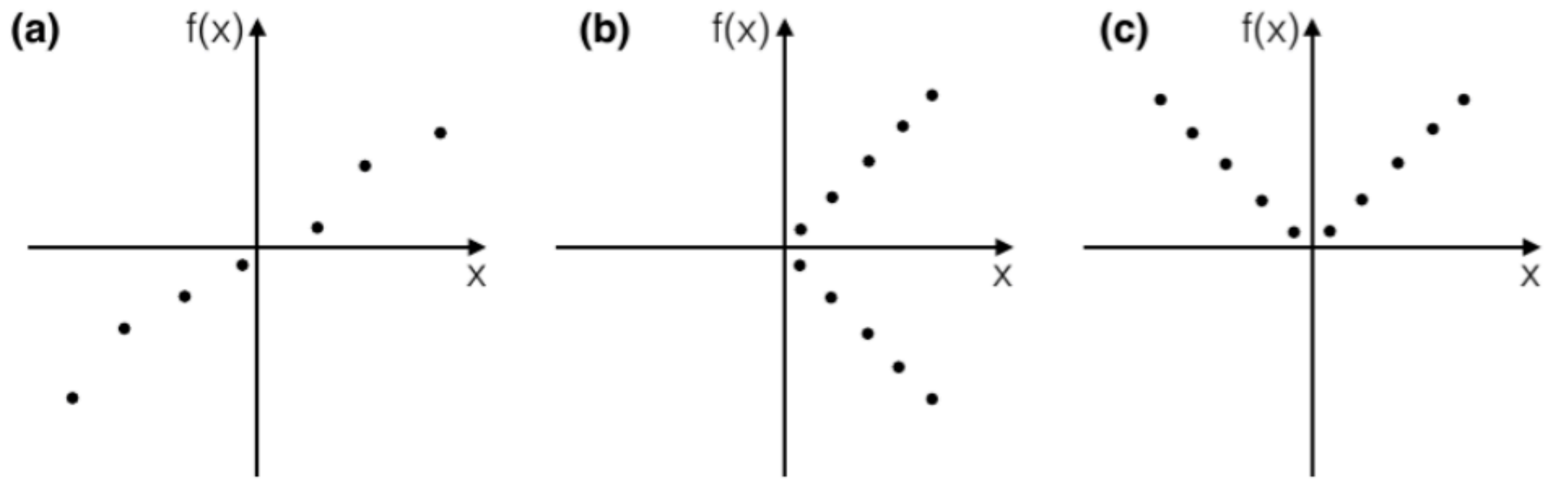
HW 1

- Compute the closed form solution to $\arg\min_{\vec{w}} \|X\vec{w} - \vec{y}\|_2^2$

- Tipps:
  - $\|\vec{x}\|_2^2 = \vec{x}^T\vec{x}$
  - $(A + B)^T = A^T + B^T$
  - $(AB)^T = B^T A^T$
  - $\frac{d}{dx} A\vec{x} = A^T$
  - $\frac{d}{dx} \vec{x}^T A = A$
  - $\frac{d}{dx} \vec{x}^T\vec{x} = 2\vec{x}$
  - $\frac{d}{dx} \vec{x}^T A\vec{x} = A\vec{x} + A^T\vec{x}$

HW 2

- Draw an approximate fit of a LLS fit of(1) a degree 1 polynomial and (2) a degree 2 polynomial



HW 3

- Assume we have perfectly generated data from $x_i + \widetilde{w}_1 = y_i$ if $i = 1, \dots, N\ i \neq j$; and **one** “outlier” $x_j + \widetilde{w}_1 + \epsilon_j = y_j$
- Write down the LLS solution (find $x + w_1 = y$) for the above problem, the solution should only depend on $\widetilde{w}_1$, ϵ_j, and N

HW 4

- Given the dataset $D_3 = \left\{([1,\ 0,\ 0], 2), \left(\left[0,\ \frac{1}{2},\ 0\right], 5\right)\right\}$
  - Compute the rank – is a pseudo-inverse required?
  - Decompose the matrix as $X = U\Sigma V^T$ (Tipp: U and V are identities – what dimensions?)
  - Write down the pseudo inverse
  - Compute $\vec{w}^*$