Machine Learning and Computational Statistics

Theodoros Georgiopoulos

Homework 1

Exercises for Unit 1: General concepts and Problem formulation

Solution for exercise 1

(a) Consider the parametric set of the quadratic functions $f_\theta : \mathbb{R} \to \mathbb{R}$. That is, for a given $\boldsymbol{x} = [x_1]^T \in \mathbb{R}$ it is

$$f_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2.$$

For $\boldsymbol{\theta} = [1, 2, 3]^T$ we have $f_{\theta_1}(x) = 1 + 2x_1 + 3x_2$.

(b) Consider the parametric set of the 3rd degree polynomials $f_\theta : \mathbb{R}^2 \to \mathbb{R}$. That is, for a given $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}$ it is

$$f_\theta(x) = \theta_0 + \theta_1 x_1^3 + \theta_2 x_2^3 + \theta_3 x_1^2 x_2 + \theta_4 x_1 x_2^2 + \theta_5 x_1^2 + \theta_6 x_2^2 + \theta_7 x_1 x_2 + \theta_8 x_1 + \theta_9 x_2.$$

For $\boldsymbol{\theta} = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]^T$ we have $f_{\theta_1} = 1 + 2x_1^3 + 3x_2^3 + 4x_1^2 x_2 + 5x_1 x_2^2 + 6x_1^2 + 7x_2^2 + 8x_1 x_2 + 9x_1 + 10x_2$.

(c) Consider the parametric set of the 3rd degree polynomials $f_\theta : \mathbb{R}^3 \to \mathbb{R}$. That is, for a given $\boldsymbol{x} = [x_1, x_2, x_3]^T \in \mathbb{R}$ it is

$$\begin{aligned} f_\theta(x) = \theta_0 &+ \theta_1 x_1^3 + \theta_2 x_2^3 + \theta_3 x_3^3 + \theta_4 x_1^2 x_2 + \theta_5 x_1 x_2^2 + \theta_6 x_1^2 x_3 + \theta_7 x_1 x_3^2 + \theta_8 x_2^2 x3 + \theta_9 x_2 x_3^2 + \\ &\theta_{10} x_3^2 x_1 + \theta_{11} x_1 x_2 x_3 + \theta_{12} x_1^2 + \theta_{13} x_2^2 + \theta_{14} x_3^2 + \theta_{15} x_1 x_2 + \theta_{16} x_2 x_3 + \theta_{17} x_1 x_3 + \theta_{18} x_1 + \\ &\theta_{19} x_2 + \theta_{20} x_3. \end{aligned}$$

Solution for exercise 2

Let $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_l]^T$ and $\boldsymbol{x} = [x_1, x_2, \dots, x_l]^T$. We have

$$(\boldsymbol{\theta^T x})\boldsymbol{x} = \left(\sum_{i=1}^{l} \theta_i x_i\right) [x_1, x_2, \dots, x_l]^T = \begin{bmatrix} x_1 \left(\sum_{i=1}^{l} \theta_i x_i\right) \\ x_2 \left(\sum_{i=1}^{l} \theta_i x_i\right) \\ \vdots \\ x_l \left(\sum_{i=1}^{l} \theta_i x_i\right) \end{bmatrix}. \tag{1}$$

Moreover,

$$(\boldsymbol{x x^T})\boldsymbol{\theta} = \begin{bmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_l \\ x_2 x_1 & x_2^2 & \dots & x_2 x_l \\ \vdots & & & \\ x_l x_1 & x_l x_2 & \dots & x_l^2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_l \end{bmatrix} = \begin{bmatrix} x_1^2\theta_1 + x_1 x_2 \theta_2 + \dots + x_1 x_l \theta_l \\ x_2 x_1 \theta_1 + x_2^2 \theta_2 + \dots + x_2 x_l \theta_l \\ \vdots \\ x_l x_1 \theta_1 + x_l x_2 \theta_2 + \dots + x_l^2 \theta_l \end{bmatrix} \tag{2}$$

From $(1), (2)$, we have

$$(\boldsymbol{\theta^T x})\boldsymbol{x} = (\boldsymbol{x x^T})\boldsymbol{\theta}.$$

Solution for exercise 3

Let $\boldsymbol{X} = [\boldsymbol{x_1^T}, \boldsymbol{x_2^T}, \dots, \boldsymbol{x_N^T}]^T$ and $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^T$. We have

$$\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{x_1 x_1^T} + \boldsymbol{x_2 x_2^T} + \dots + \boldsymbol{x_n x_n^T} = \sum_{i=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^T.$$

Additionally,

$$\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{x_1} y_1 + \boldsymbol{x_2} y_2 + \dots + \boldsymbol{x_n} y_n = \sum_{i=1}^{N} y_n \boldsymbol{x}_n.$$

Solution for exercise 4

We have one dataset $\{(x_i, y_i), x_i \in R, y_i \in R, i = 1, \dots, 5\}$ with the values $\{(2, 2.01), (4, 4.01), (-2, -2.01), (-3, -3.01), (-1, -1.01)\}$. So, $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & -2 \\ 1 & -3 \\ 1 & -1 \end{bmatrix}$,

$\boldsymbol{y} = \begin{bmatrix} 2.01 \\ 4.01 \\ -2.01 \\ -3.01 \\ -1.01 \end{bmatrix}$ and we want to find $\boldsymbol{\theta} = [\theta_0, \theta_1]$. The Least Squares (LS) estimator, which minimizes the mean of square error, gives

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}.$$

So, $\boldsymbol{\theta} = [-0.002, 1.003]$.