Statistik un Data Science für die Informatik

Humboldt-Universität zu Berlin, SS 22, Prof. Dr. Alan Akbik

Bearbeitung von Übungsblatt 5

Namen und Matrikelnummer der Abgebenden:

Gruppe AG-24	Name	CMS-Benutzername	Matrikelnr
	Kristina Pianykh	pianykhk	617331
	Miguel Nuno Carqueijeiro Athouguia	carqueim	618203
	Winston Winston	winstonw	602307

Aufgabe 1

Gegeben: x $= \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} \in \mathbb{R}^{5x1}$, y $= \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \\ 6 \end{pmatrix} \in \mathbb{R}^{5x1}$ and $f(x) = a + bx$.

To be found: a and b.

Let W be a weight matrix such that w $= \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^{2x1}$

The loss function (i.e. the mean squared error (MSE) function) is defined as follows:

$L(w) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i) = \frac{1}{n}||Xw - y||^2$.

The MSE function is continuous and differentiable. It also has one minimum, the global minimum. To find the solution to this equation (namely, to find the weights vector w with the least mean squared error), we need to compute the derivative of the loss function and set it to 0:

$$\frac{dL}{dw} = \begin{pmatrix} \frac{dL}{dw_0} \\ \frac{dL}{dw_1} \end{pmatrix} = 0$$

The closed-form solution:

$$\frac{dL}{dw} = \begin{pmatrix} \frac{dL}{dw_0} \\ \frac{dL}{dw_1} \end{pmatrix} = X^T Xw - X^T y = 0 \quad | + X^T y$$

$$\Longleftrightarrow X^T Xw = X^T y \quad | * (X^T X)^{-1}$$

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

X is a design matrix (meaning it has an additional column of 1s to the original x vector to fully vectorize the computations): $f(x) = a + bx = ax_1 + bx_2 = a * 1 + b * x_2$.

$$X = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \in \mathbb{R}^{5x2}, X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \in \mathbb{R}^{2x5}$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix} \in \mathbb{R}^{2x2}$$

To compute the inverse of $X^T X$ we need to apply Cramer's rule:

$$(X^T X)^{-1} = \frac{1}{det(X^T X)} \begin{pmatrix} (-1)^{1+1} det(X^T X_{11)} & (-1)^{2+1} det(X^T X_{21)} \\ (-1)^{1+2} det(X^T X_{12)} & (-1)^{2+2} det(X^T X_{22)} \end{pmatrix} =$$

$$= \frac{1}{det(X^T X)} \begin{pmatrix} det(X^T X_{11)} & -det(X^T X_{21)} \\ -det(X^T X_{12)} & det(X^T X_{22)} \end{pmatrix}$$

$$(X^T X)^{-1} = \frac{1}{50} \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix} \in \mathbb{R}^{2x2}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \end{pmatrix} \in \mathbb{R}^{2x1}$$

$$w = (X^T X)^{-1} X^T y = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} 10 \\ 10 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in \mathbb{R}^{2x1}$$

$\implies w_0 = a = 2, w_1 = b = 1$

$\implies f(x) = 2 + x.$