"Machine Learning and Computational Statistics"

1st Homework

Exercise 1:

Let $\boldsymbol{x} = [x_1, x_2]^T$ be the input/feature vector and y the output variable in a regression task. Consider the following scenarios that relate y with $\boldsymbol{x}$:

(i) $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + {x_1}^2$.

(ii) $y = sign(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 {x_1}^2)$

(iii) $y = 2x_1 + sign(3 - 7)x_2 + ReLU(3)x_1 x_2$

(iv) $y = \theta + \theta x_1 + \theta x_2 + \theta x_1 x_2$

(a) How many and which are the parameters involved in each scenario?

(b) In which of the above scenarios y has linear dependence with the involved parameters?

Note: It is $sign(z) = \begin{cases} 1, & z \geq 0 \\ 0, & otherwise \end{cases}$ and $ReLU(z) = \begin{cases} z, & z \geq 0 \\ 0, & otherwise \end{cases}$.

Exercise 2:

Consider different Machine Learning tasks where the input $\boldsymbol{x}$ is two-dimensional, i.e. $\boldsymbol{x} = [x_1, x_2]^T$, and the output y is scalar. Assume also that in each case, a data set that consists of pairs $(y_i, \boldsymbol{x}_i) \equiv (y_i, [x_{i1}, x_{i2}]^T)$, $i = 1, \dots, N$, is available. Which of the following models are parametric and which are non-parametric?

1. $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

2. $y = min(x_1, x_2)$

3. $y = ReLU(\theta_0 + \theta_1 x_1)$

4. $y = \sum_{i=1}^{N} \theta_i (x_{i1} x_1 - x_{i2} x_2)$

Note: It is $min(x_1, x_2) = \begin{cases} x_1, & x_1 \geq x_2 \\ x_2, & otherwise \end{cases}$ and $ReLU$ is defined as in exercise 1.

Exercise 3:

(a) Define the parametric set of the **quadratic** functions $f_{\boldsymbol{\theta}}: R \to R$ and give two instances of it. What is the dimensionality of $\boldsymbol{\theta}$?

(b) Define the parametric set of the **3rd degree polynomials** $f_{\boldsymbol{\theta}}: R^2 \to R$ and give two instances of it. What is the dimensionality of $\boldsymbol{\theta}$?

(c) Define the parametric set of the **3rd degree polynomials** $f_{\boldsymbol{\theta}}: R^3 \to R$ and give two instances of it. What is the dimensionality of $\boldsymbol{\theta}$?

(d) Consider the function $f_{\boldsymbol{\theta}}(\boldsymbol{x}): R^5 \to R$, $f_{\theta}(\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{\theta}^T\boldsymbol{x})}$. Define the associated parametric set and give two instances of it. What is the dimensionality of $\boldsymbol{\theta}$?

(e) In which of the above cases $f_{\boldsymbol{\theta}}$ is linear with respect to $\boldsymbol{\theta}$?

Exercise 4:

Verify that for two l-dimensional column vectors $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_l]^T$ and $\boldsymbol{x} = [x_1, x_2, \dots, x_l]^T$ it holds: $(\boldsymbol{\theta}^T\boldsymbol{x})\boldsymbol{x} = (\boldsymbol{x}\,\boldsymbol{x}^T)\boldsymbol{\theta}$.

Hint: (a) Take the left part of the equation, substitute $\boldsymbol{\theta}$ and $\boldsymbol{x}$ and perform algebraic operations, in order to bring it in its reduced form. Work similarly with the right part and compare the reduced forms of the two parts of the equation.

(b) The **inner product** of two (real-valued) column vectors $\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_l \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_l \end{bmatrix}$ is computed as $\boldsymbol{u}^T\boldsymbol{v} = [u_1, u_2, \dots, u_l] \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_l \end{bmatrix} = u_1v_1 + \cdots + u_lv_l = \sum_{i=1}^{l} u_iv_i$

Exercise 5:

Consider the vectors $\boldsymbol{x}_n = [x_{n1}, x_{n2}, \dots, x_{nl}]^T, n = 1, \dots, N$. Define the $N \times l$ matrix X and the N-dimensional column vector $\boldsymbol{y}$ as follows:

$$X = \begin{bmatrix} \boldsymbol{x}_1{}^T \\ \boldsymbol{x}_2{}^T \\ \vdots \\ \boldsymbol{x}_N{}^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nl} \end{bmatrix} \text{ and } \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

(Note that the rows of X are the vectors $\boldsymbol{x}_n, n = 1, \dots, N$).

(a) Verify the following identities:

$X^T X = \sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^T$ and $X^T \boldsymbol{y} = \sum_{n=1}^{N} y_n \boldsymbol{x}_n$.

(b) What is the size of X, $\boldsymbol{y}$, $X^T X$ and $X^T \boldsymbol{y}$?

(c) Assume that a column vector of 1's is added in front of the column $\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{N1} \end{bmatrix}$ of matrix X.

(c-i) What will be the changes in the dimensionality of the quantities in (b)?

(c-ii) Do the identities given in (a) still hold?

<u>Exercise 6:</u> A body moves on a straight line and performs a smoothly accelerating motion (we begin to study its motion at the time instance $t = 0$). In the following table is given the velocity at certain time instances

t (sec)	1	2	3	4	5
v (m/sec)	5.1	6.8	9.2	10.9	13.1

(a) Estimate the initial velocity and the acceleration of the body, based on the above measurements, utilizing the least squares error criterion.
(b) Write down the equation that expresses the velocity of the body as a function of time t.
(c) Estimate the velocity of the body at $t = 2.3$.

<u>Hints:</u>

(i) The velocity v of a body moving on a straight line and performing a smoothly accelerating motion is given by the equation

$$v = v_0 + a \cdot t$$

where t is the time instance, a is the acceleration and v_0 the initial speed (at time instance $t = 0$).

(ii) The previous table of values is associated with the following data set

$$\{(y_i, x_i), i = 1, \ldots, 5\} \equiv \{(v_i, t_i), i = 1, \ldots, 5\}$$
$$= \{(5.1, 1), (6.8, 2), (9.2, 3), (10.9, 4), (13.1, 5)\}$$

(iii) Define $\boldsymbol{\theta} = [v_0, a]^T$, construct the matrix X and the vector $\boldsymbol{y}$, and utilize the equation that gives the Least squares estimation (slide 50 of the 1st lecture) to estimate its value.

(iv) The inverse of a $2x2$ matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $A^{-1} = \frac{1}{D}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, where $D = a \cdot d - b \cdot c$.