Homework 1
Giagtos Stylianos
f3352410

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

i) $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + x_1^2$
ii) $y = \text{sign}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2)$
iii) $y = 2x_1 + \text{sign}(3-7) x_2 + \text{RELU}(3) x_1 x_2$
iv) $y = \theta + \theta x_1 + \theta x_2 + \theta x_1 x_2$

α) 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?

• Scenario i) $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + x_1^2$
a) There are 4 parameters $\theta_0, \theta_1, \theta_2, \theta_3$
b) y has linear dependence on the parameters $\theta_0, \theta_1, \theta_2, \theta_3$ as each term appears in linear form with respect to parameters. Although the dependence regarding the parameters is linear, due to the term x_1^2 y is not linear to the input x's.

• Scenario ii) $y = \text{sign}(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2)$
a) There are 5 parameters $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4$
b) y is not linearly dependent on the parameters as, sign function is not a linear function

• Scenario iii) $y = 2x_1 + \text{sign}(3-7) x_2 + \text{ReLU}(3) x_1 x_2$
α) There are no parameters.
b) y has no linear dependence on parameters as there are no parameters.

• Scenario iv) $y = \theta + \theta x_1 + \theta x_2 + \theta x_1 x_2$

a) The parameter here is one and it is θ.

b) y is linearly dependent on θ

Exercise 2.

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

- Model 1: $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$
  This model is parametric including parameters $\theta_0, \theta_1, \theta_2$
- Model 2: $y = \min(x_1, x_2)$
  This model is non-paremetric as it depends only on the x_1, x_2 inputs.
- Model 3: $\mathrm{ReLU}(\theta_0 + \theta_1 x_1)$
  This model is paremetric as it contains parameters θ_0, θ_1
- Model 4: $\sum_{i=1}^{N} \theta_i (x_{i1} x_1 - x_{i2} x_2)$
  This model is parametric as it contains θ_i parameters. depending on the summation.

Exercise 3.

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

The function has the following form:

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

Instances are: $f_\theta(x) = 5 + 6x + 7x^2$

$f_\theta(x) = -3 + 2x + 8x^2$

The dimensionality of θ is 3 ($\theta_0, \theta_1, \theta_2$)

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

The function has the following form:

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

Instances are:

$$f_\theta(x_1, x_2) = 1 + 2x_1 + 3x_2 + x_1^2 + x_1 x_2 + x_2^2 + x_1^3 + x_1^2 x_2 + x_1 x_2^2 + x_2^3$$

$$f_\theta(x_1, x_2) = 8 + 12x_1 + 3x_2 + x_1^2 + 7x_1 x_2 + x_2^2 + 2x_1^3 + 6x_1^2 x_2 + 4x_1 x_2^2 + 2x_2^3$$

The dimensionality of θ is 10. (θ_0 to θ_9)

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

The function has the following form:

$f_\theta(x_1, x_2, x_3) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_1^2 + \theta_5 x_2 x_3 + \dots + \theta_{19} x_3^3$

Instances are:

$f_\theta(x_1, x_2 x_3) = 1 + x_1 + 2x_2 + 3x_3 + x_1^2 + 5x_2 x_3 + 7x_1 x_2^2 + \dots + 10x_3^3$

$f_\theta(x_1, x_2, x_3) = 4 + x_1 + x_2 + x_3 + \dots + x_3^3.$

The dimensionality of θ is 20: (θ_0 to θ_{19})

d) Consider the function $f_\theta(x): \mathbb{R}^5 \to \mathbb{R}, f_\theta(x) = \dfrac{1}{1+\exp(-\theta^T x)}$.

Define the associated parametric set and give two instances of it. What is the dimensionality of θ?

Parametric Set: consists of all functions of the form,

$$f_\theta(x) = \frac{1}{1+\exp(-\theta_0 - \theta_1 x_1 - \theta_2 x_2 - \theta_3 x_3 - \theta_4 x_4 - \theta_5 x_5)}$$

Two instances are:

$$f_\theta(x) = \frac{1}{1+\exp(-(1+2x_1+3x_2+4x_3+5x_4+6x_5))}$$

Where $\theta = [1, 2, 3, 4, 5, 6]$

$$f_\theta(x) = \frac{1}{1+\exp(-(0.3+2x_1+0.1x_2+0.7x_3+0.8x_4+1.2x_5))}.$$

Where θ is $[0,3, 2, 0.1, 0.7, 0.8, 1,2]$

The dimensionality of θ is 6: $[\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$

e) In which of the above cases f_θ is linear with respect to θ?

f_θ is linear with respect to θ in cases

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

b) $f_\theta(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \ldots + \theta_9 x^2$

c) $f_\theta(x_1, x_2, x_3) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots + \theta_{19} x_3^3$

In all above cases the function f_θ is linear with respect to θ as each θ parameter appears linearly despite products and powers of xs.

Exercise 4

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

a) Left part of the equation is: $(\theta^T x)x$

The product $\theta^T x$ is computed as

$$\theta^T x=\theta_1 x_1+\theta_2 x_2+\dots+\theta_l x_l=\sum_{i=1}^{l}\theta_i x_i$$

$$(\theta^T x)x=\left(\sum_{i=1}^{l}\theta_i x_i\right)\begin{bmatrix}x_1\\x_2\\\vdots\\x_l\end{bmatrix}=\begin{bmatrix}(\sum_{i=1}^{l}\theta_i x_i)x_1\\(\sum_{i=1}^{l}\theta_i x_i)x_2\\\vdots\\(\sum_{i=1}^{l}\theta_i x_i)x_l\end{bmatrix}$$

b) The right part of the equation is: $(xx^T)\theta$:

$$xx^T=\begin{bmatrix}x_1\\x_2\\\vdots\\x_l\end{bmatrix}\begin{bmatrix}x_1 & x_2 & \dots & x_l\end{bmatrix}=\begin{bmatrix}x_1^2 & x_1x_2 & \dots & x_1x_l\\x_1x_2 & x_2^2 & \dots & x_2x_l\\\vdots & \vdots & \ddots & \vdots\\x_lx_1 & x_lx_2 & \dots & x_l^2\end{bmatrix}$$

$$(xx^T)\cdot\theta=\begin{bmatrix}x_1^2 & x_1x_2 & \dots & x_1x_l\\x_1x_2 & x_2^2 & \dots & x_2x_l\\\vdots & \vdots & \ddots & \vdots\\x_lx_1 & x_lx_2 & \dots & x_l^2\end{bmatrix}\cdot\begin{bmatrix}\theta_1\\\theta_2\\\vdots\\\theta_l\end{bmatrix}=\begin{bmatrix}x_1(x_1\theta_1+\dots+x_l\theta_l)\\x_2(x_1\theta_1+\dots+x_l\theta_l)\\\vdots\\x_l(x_1\theta_1+\dots+x_l\theta_l)\end{bmatrix}$$

$$=\begin{bmatrix}x_1\left(\sum_{i=1}^{l}(x_i\theta_i)\right)\\x_2\left(\sum_{i=1}^{l}(x_i\theta_i)\right)\\\vdots\\x_l\left(\sum_{i=1}^{l}(x_i\theta_i)\right)\end{bmatrix}=(\theta^T x)\cdot x.$$

So we proved that $(\theta^T x)\cdot x=(x\cdot x^T)\theta$.

Exercise 5

a) Verify the following identities
$X^T X = \sum_{u=1}^{N} X_u X_u^T$ and $X^T y = \sum_{u=1}^{N} y_u X_u$

i) $X^T X = \sum_{u=1}^{N} X_u X_u^T$

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_N^T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1l} \\ X_{21} & X_{22} & \cdots & X_{2l} \\ \vdots & \vdots & & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Nl} \end{bmatrix}$$

$$X^T X = \begin{bmatrix} \sum_{u=1}^{N} X_{u1}^2 & \sum_{u=1}^{N} X_{u1} X_{u2} & \cdots & \sum_{u=1}^{N} X_{u1} X_{ul} \\ \sum_{u=1}^{N} X_{u2} X_{u1} & \sum_{u=1}^{N} X_{u2}^2 & \cdots & \sum_{u=1}^{N} X_{u2} X_{ul} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{u=1}^{N} X_{ul} X_{ul} & \sum_{u=1}^{N} X_{ul} X_{u2} & \cdots & \sum_{u=1}^{N} X_{ul}^2 \end{bmatrix}$$

So, $X^T X = \sum_{u=1}^{N} X_u X_u^T$

ii) $X^T y = \sum_{u=1}^{N} y_u X_u$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X^T y = \begin{bmatrix} \sum_{u=1}^{N} X_{u1} y_n \\ \sum_{u=1}^{N} X_{u2} y_u \\ \vdots \\ \sum_{u=1}^{N} X_{u1} y_u \end{bmatrix} = \sum_{n=1}^{N} y_n X_n$$

b) What is the size of $X, y, X^T X$ and $X^T y$

1) Size of X is $N \times l$ (samples: rows, features: columns)
2) Size of y is $N \times 1$
3) Size of $X^T X$ is $l \times l$ ($X^T: l \times N$, $X: N \times l$)
4) Size of $X^T y$ is $l \times 1$

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

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

By adding a column vector of 1's in matrix X we get the following changes in sizes:

- Matrix X has new size of $l+1$ columns therefore size of X is now $N \times (l+1)$.

- Vector y remains same.

- Matrix $X^T X$ is the product of the new X^T with dimensions $(l+1) \times N$ and the new X with $(N \times (l+1))$. So the $X^T X$ will have $(l+1) \times (l+1)$ dimensions.

- Vector $X^T y$ will be $(l+1) \times 1$ as the columns were increased by 1.

ii) Do the identities given in (α) still hold?

1. $X^T X = \sum_{n=1}^{N} x_n x_n^T$

The matrix X' is: $X' = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1l} \\ 1 & x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nl} \end{bmatrix}$

$$X'^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1l} & x_{2l} & \cdots & x_{Nl} \end{bmatrix}$$

$$X'^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1l} & x_{2l} & \cdots & x_{Nl} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1l} \\ 1 & x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nl} \end{bmatrix}$$

$$X'^T X' = \begin{bmatrix} N & \sum_{n=1}^{N} x_{n1} & \cdots & \sum_{n=1}^{N} x_{nl} \\ \sum_{n=1}^{N} x_{n1} & \sum_{n=1}^{N} x_{n1}^2 & \cdots & \sum_{n=1}^{N} x_{n1} x_{nl} \\ \sum_{n=1}^{N} x_{n2} & \sum_{n=1}^{N} x_{n1} x_{n2} & \cdots & \sum_{n=1}^{N} x_{n2} x_{nl} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{n=1}^{N} x_{nl} & \sum_{n=1}^{N} x_{n1} x_{nl} & \cdots & \sum_{n=1}^{N} x_{nl}^2 \end{bmatrix}$$

$$x'_u = \begin{bmatrix} 1 \\ x_{u1} \\ x_{u2} \\ \vdots \\ x_{up} \end{bmatrix}, \quad x'_u x'^T_u = \begin{bmatrix} 1 & x_{u1} & x_{u2} & \cdots & x_{up} \\ x_{u1} & x^2_{u1} & x_{u1}x_{u2} & \cdots & x_{u1}x_{up} \\ x_{u2} & x_{u1}x_{u2} & x^2_{u2} & \cdots & x_{u2}x_{up} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{up} & x_{u1}x_{up} & x_{u2}x_{up} & & x^2_{up} \end{bmatrix}$$

$$\sum_{u=1}^{N} x'_u x'^T_u = \begin{bmatrix} N & \sum_{u=1}^{N} x_{u1} & \cdots & \sum_{u=1}^{N} x_{up} \\ \sum_{u=1}^{N} x_{u1} & \sum_{u=1}^{N} x^2_{u1} & \cdots & \sum_{u=1}^{N} x_{u1}x_{up} \\ \sum_{u=1}^{N} x_{u2} & \sum_{u=1}^{N} x_{u1}x_{u2} & \cdots & \sum_{u=1}^{N} x_{u2}x_{up} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{u=1}^{N} x_{up} & \sum_{u=1}^{N} x_{u1}x_{up} & \cdots & \sum_{u=1}^{N} x^2_{up} \end{bmatrix}$$

So it still holds.

2) $x'^T y = \sum_{u=1}^{N} y_u x'_u$

$$x'^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{Np} \end{bmatrix}, \quad x'^T y = \begin{bmatrix} \sum_{u=1}^{N} y_u \\ \sum_{u=1}^{N} x_{u1} y_u \\ \sum_{u=1}^{N} x_{u2} y_u \\ \vdots \\ \sum_{u=1}^{N} x_{up} y_u \end{bmatrix}$$

$$x'_u = \begin{bmatrix} 1 \\ x_{u1} \\ x_{u2} \\ \vdots \\ x_{u2} \end{bmatrix} \Rightarrow \sum_{u=1}^{N} y_u x'_u = \sum_{u=1}^{N} y_u \begin{bmatrix} 1 \\ x_{u1} \\ \vdots \\ x_{up} \end{bmatrix} = \begin{bmatrix} \sum_{u=1}^{N} y_u \\ \sum_{u=1}^{N} y_u x_{u1} \\ \vdots \\ \sum_{u=1}^{N} y_u x_{up} \end{bmatrix}$$

So $x'^T y = \sum_{u=1}^{N} y_u x'_u$.

Exercise 6

A body moves on straight line and performs a smoothly accelerating motion. 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

By using the given formula $v = v_0 + at$ we get that v = velocity, v_0 = initial velocity, acceleration is a and t is time.

So we get the matrix X by adding one column with 1's for v_0 and t values in second column.

The vector y is for velocity instances.

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}, \quad y = \begin{bmatrix} 5.1 \\ 6.8 \\ 9.2 \\ 10.9 \\ 13.1 \end{bmatrix}$$

The Least Squares (LS) estimator is

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

$$X^T \cdot X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix}$$

$$X^T \cdot y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 5.1 \\ 6.8 \\ 9.2 \\ 10.9 \\ 13.1 \end{bmatrix} = \begin{bmatrix} 45.1 \\ 155.4 \end{bmatrix}$$

$(X^T \cdot X)^{-1}$: $(X^T \cdot X)^{-1} = \dfrac{1}{\det(X^T \cdot X)} \cdot adj(X^T \cdot X)$

$\det(X^T \cdot X) = 5 \cdot 55 - 15 \cdot 15 = 275 - 225 = 50.$

$adj(X^T \cdot X) = \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix}$

$(X^T \cdot X^{-1}) = \dfrac{1}{50} \begin{bmatrix} 55 & -15 \\ -15 & 5 \end{bmatrix} = \begin{bmatrix} 1.1 & -0.3 \\ -0.3 & 0.1 \end{bmatrix}$

So $\theta = (X^T \cdot X)^{-1} \cdot X^T y$

$= \begin{bmatrix} 1.1 & -0.3 \\ -0.3 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 45.1 \\ 155.4 \end{bmatrix} = \begin{bmatrix} 2.99 \\ 2.01 \end{bmatrix}$

So $\theta = \begin{bmatrix} v_0 \\ \alpha \end{bmatrix} = \begin{bmatrix} 2.99 \\ 2.01 \end{bmatrix}$

Therefore $v_0 = 2.99$ m/s

$\alpha = 2.01$ m/s^2

b) Equation of velocity of the body as function of t.

$v(t) = v_0 + \alpha t$

$v(t) = 2.99 + 2.01 t.$

c) Estimate the velocity of the body at $t = 2.3$

$v(2.3) = 2.99 + 2.01 \cdot 2.3 = 7.61$ m/s.