Logistic regression – case study:

Task:
- find the best coefficients

Input data:

datapoint	hours studying	results 1=passed 0=failed
i	**x_i**	**y_i**
0	1	0
1	2	0
2	3	0
3	4	0
4	5	1
5	6	0
6	7	1
7	8	1
8	9	1
9	10	1

Sigmoid function:

$$z_i = \theta_0 . 1 + \theta_j . x_{ij} \tag{1}$$

$$\sigma_{z_i} = \frac{1}{1 + e^{-z_i}} \tag{2}$$

Cost function:
- derivative of the cost function

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[y_i . log(\sigma_{zi}) + (1 - y_i) . log(1 - \sigma_{zi})]$$

.

. (3)

$$\frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{n}\sum_{i=1}^{n}[(y_i - \sigma_{zi}) . x_{ij}]$$

First iteration:

- assume starting values:

$\theta_0 = 0, \theta_1 = 0$

- calculate sigmoid function

$$
\begin{gathered}
z_i = \theta_0 .1 + \theta_j . x_{ij} \\
z_0 = \theta_0 .1 + \theta_1 . x_{01} = 0 * 1 + 0 * 1 = 0 \\
. \\
. \\
. \\
z_9 = \theta_0 .1 + \theta_1 . x_{91} = 0 * 1 + 0 * 10 = 0
\end{gathered} \tag{4}
$$

$$
\begin{gathered}
\sigma_{z_i} = \frac{1}{1 + e^{-z_i}} \\
\sigma_{z_0} = \frac{1}{1 + e^{-z_0}} = \frac{1}{1 + e^{-0}} = 0.5 \\
. \\
. \\
. \\
\sigma_{z_9} = \frac{1}{1 + e^{-z_9}} = \frac{1}{1 + e^{-0}} = 0.5
\end{gathered} \tag{5}
$$

- cost function of the first variable

$$
\begin{gathered}
\frac{\partial J(\theta)}{\partial \theta_0} = -\frac{1}{n} \sum_{i=1}^{n} [(y_i - \sigma_{zi}) . x_{i0}] = -\frac{1}{n} \sum_{i=1}^{n} (y_i - \sigma_{zi}) \\
\frac{\partial J(\theta)}{\partial \theta_0} = -\frac{1}{9} [(y_0 - \sigma_{z0}) + (y_1 - \sigma_{z1}) + \cdots + (y_9 - \sigma_{z9})]
\end{gathered} \tag{6}
$$

- cost function of the second variable

$$
\begin{gathered}
\frac{\partial J(\theta)}{\partial \theta_1} = -\frac{1}{n} \sum_{i=1}^{n} [(y_i - \sigma_{zi}) . x_{i1}] \\
\frac{\partial J(\theta)}{\partial \theta_1} = -\frac{1}{9} [(y_0 - \sigma_{z0}) x_{01} + (y_1 - \sigma_{z1}) x_{11} + \cdots + (y_9 - \sigma_{z9}) x_{91}]
\end{gathered} \tag{7}
$$

- update Gradient descent (**:=**)

$$\theta_0 := \theta_0 - \alpha \frac{\partial J(\theta)}{\partial \theta_0}$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial J(\theta)}{\partial \theta_1} \tag{8}$$

- repeat the cycle with updated coefficients $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta_1}$ until convergence (set the threshold $\boldsymbol{\varepsilon}$)

$$\left|J(\theta^{(k)}) - J(\theta^{(k-1)})\right| < \epsilon$$

or

$$\left|\theta_0^{(t+1)} - \theta_0^{(t)}\right| < \epsilon \;\; and \;\; \left|\theta_1^{(t+1)} - \theta_1^{(t)}\right| < \epsilon \tag{8}$$

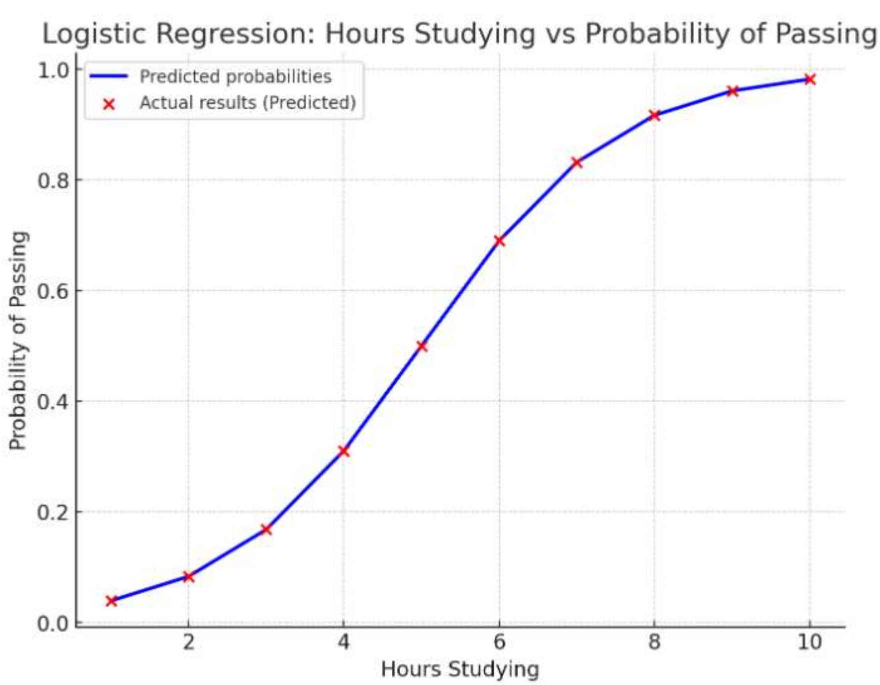
or

$$\left|\frac{\partial J(\theta)}{\partial \theta_0}\right| < \epsilon \;\; and \;\; \left|\frac{\partial J(\theta)}{\partial \theta_1}\right| < \epsilon$$

Plotted results:

- final sigmoid function represent the probability that student will fail or not

$$\sigma_i = \frac{1}{1 + e^{(-\theta_0 . + \theta_1 . x_i)}} \tag{11}$$



Iterations using matrix-vector operations:

- capital letters => **matrices**
- lower case letters => **vectors**

- sigmoid function
- assumption:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$z_i = \theta_0 . 1 + \theta_j . x_{ij}$$

$$\boldsymbol{z} = \boldsymbol{X}.\boldsymbol{\theta}$$

$$\boldsymbol{z} = \boldsymbol{X}.\boldsymbol{\theta} = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \\ \vdots & \vdots \\ x_{90} & x_{91} \end{bmatrix} . \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} \theta_0 . x_{00} + \theta_1 x_{01} \\ \theta_0 . x_{10} + \theta_1 x_{11} \\ \vdots \\ \theta_0 . x_{90} + \theta_1 x_{91} \end{bmatrix} = \begin{bmatrix} 0 * 1 + 0 * 1 \\ 0 * 1 + 0 * 2 \\ \vdots \\ 0 * 1 + 0 * 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9}$$

$$\sigma_{z_i} = \frac{1}{1 + e^{-z_i}}$$

$$\boldsymbol{P} = \frac{1}{1 + e^{-\boldsymbol{z}}} = \begin{bmatrix} 0.5 \\ 0.5 \\ \vdots \\ 0.5 \end{bmatrix} \tag{10}$$

- cost function

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{n} \sum_{i=1}^{n} [(y_i - \sigma_{zi}). x_{ij}]$$

$$\frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{n} \boldsymbol{X}^T (\boldsymbol{y} - \boldsymbol{P})$$

$$\boldsymbol{y} - \boldsymbol{P} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_9 \end{bmatrix} - \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \\ \vdots \\ 0.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -0.5 \\ \vdots \\ 0.5 \end{bmatrix} \tag{11}$$

$$\boldsymbol{X}^T (\boldsymbol{y} - \boldsymbol{P}) = \begin{bmatrix} 1 & 1 & . & . & 1 \\ 1 & 2 & . & . & 10 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \\ \vdots \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 * 0.5 + 1 * 0.5 + \cdots + 1 * 0.5 \\ 1 * 0.5 + 2 * 0.5 + \cdots + 10 * 0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 11.5 \end{bmatrix}$$

- update coefficients

$$\theta := \boldsymbol{\theta} - \alpha \frac{1}{n} \boldsymbol{X}^T (\boldsymbol{y} - \boldsymbol{P})$$

$$\theta := \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 \frac{1}{10} \begin{bmatrix} 0 \\ 11.5 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.115 \end{bmatrix} \tag{11}$$