Linear Regression with mean squared error

0.1 Method

The linear regression aims at modelling relationships between a dependant variable and a set of independent variables.

It takes as input a vector $\mathbf{x} \in \mathbb{R}^{m*n}$ and predicts the value of $\mathbf{y} \in \mathbb{R}^{m}$.

The regression line for n feature is represented as:

$$\hat{y}(x_i) = b_0 + \sum_{j=1}^{n} b_j * x_{i,j} + \epsilon_i \tag{1}$$

where:

- $\hat{y}(x_i)$ is the predicted value;
- $b_0, ..., b_n$ are the coefficients;
- ϵ_i is the residual error;

We can write this model, using matrices, as follow:

$$\mathbf{y} = \mathbf{bX} + \epsilon \tag{2}$$

where:

- $\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix}$
- $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$ $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

In the next section I'm going to compute the mean squared error without taking into consideration b_0 and ϵ.

0.2 Minimising Mean Squared Error

If $\hat{\mathbf{y}}$ is the vector of predicted values and $\mathbf{y}$ is the real vector, we can measure the performance of the model with the mean squared error, which is equal to

$$MSE = \frac{1}{m} \|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 \tag{3}$$

To minimise it, we can compute its gradient and solve for where it's zero:

Dimostrazione. $\nabla_{\mathbf{b}} \frac{1}{m} \|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 = 0$
$\nabla_{\mathbf{b}} (\mathbf{X}\mathbf{b} - \mathbf{y})^T (\mathbf{X}\mathbf{b} - \mathbf{y}) = 0$
$\nabla_{\mathbf{b}} (\mathbf{b}^T \mathbf{X}^T \mathbf{X}\mathbf{b} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{y}^T \mathbf{y}) = 0$
$2\mathbf{X}^T \mathbf{X}\mathbf{b} - 2\mathbf{X}^T \mathbf{y} = 0$
$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$

□