KAZAKH-BRITISH TECHNICAL UNIVERSITY

Introduction to Machine Learning
Week 5

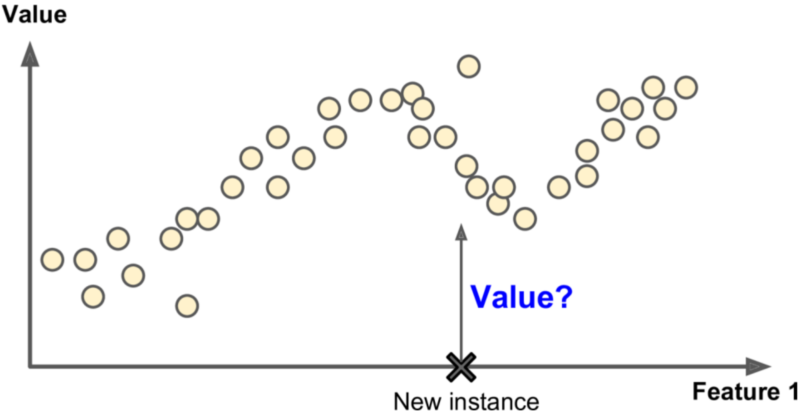
Olivier JAYLET

School of Information Technology and Engineering

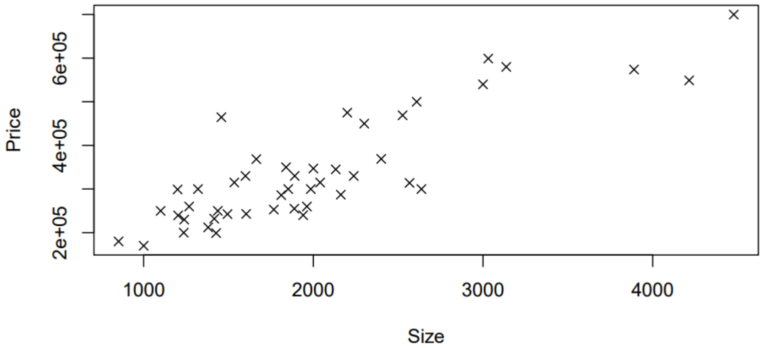
Context

- We talk about **supervised learning** when we know the “true response” for each observation.
- A **regression** problem consists in predicting a real value...
- ...unlike a **classification** problem which consists in predicting a discrete value.

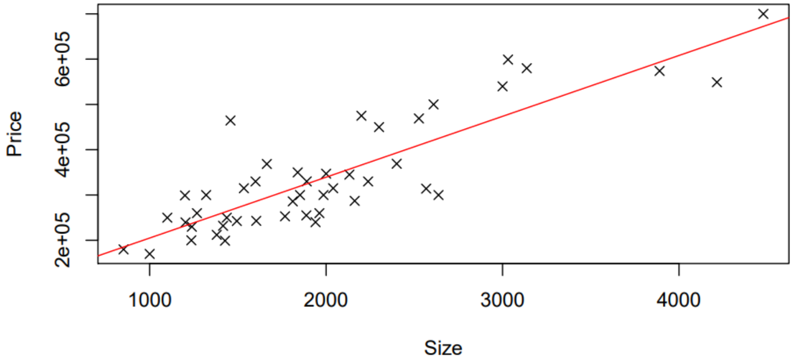
Context



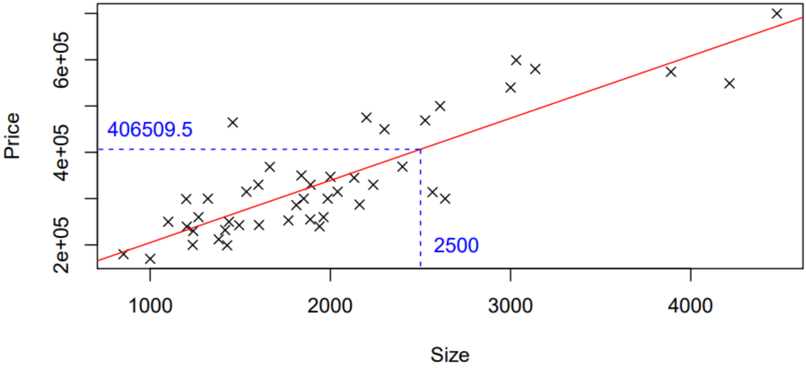
Example



Regression Line



Prediction



Learning sample

x = Size	y = Price
1600	329900
2400	369000
1416	232000
3000	539900
1985	299900
1534	314900

Notations:

- n: number of observations in the learning sample
- x: explanatory variable (or predictor)
- y: variable to explain (or target)
- (x, y): one observation (or example)
- $(x^{(i)}, y^{(i)})$: the i-th observation

Objective

Using this dataset, we will "fit", "train" or "estimate" a model, such that with a new sample x, we could "predict" y.

x = Size	**y = Price**
1600	329900
2400	369000
1416	232000
3000	539900
1985	299900
1534	314900
3150	?
4840	?
5170	?

Model definition

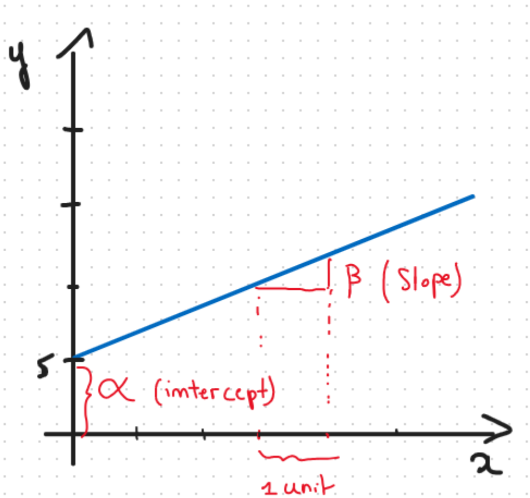
The univariate regression is defined as :

$$y_i = \alpha + \beta x_i + \epsilon$$

Where:

- y_i is the dependent variable,
- x_i is the independent variable,
- α is the intercept (constant) of the regression line,
- β is the slope of the regression line,
- ϵ is the error term (or unpredictable random disturbance).

Representation



Parameters interpretation

α : Value of the dependent variable (y) when the independent variable (x) is zero.
β : Tells how much y increases (or decreases) when the variable x increases by one unit.

Linear regression

We want to find (estimate) the parameters $\hat{\alpha}$ & $\hat{\beta}$, such that the predicted values $\hat{y} = \hat{\alpha} + \hat{\beta}x$ are as close as possible to the actual values y,
i.e. minimize a cost function (quantified difference between y & $\hat{y}$).

Residuals definition

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{1}$$

The residual ϵ_i is the difference between the observed value y_i and the predicted value $\hat{y}_i$:

$$\epsilon_i = y_i - (\alpha + \beta x_i) \tag{2}$$

Substituting $\hat{y}_i = \alpha + \beta x_i$, we get:

$$\epsilon_i = y_i - \hat{y}_i \tag{3}$$

Sum of Squared Residuals (SSR)

The Sum of Squared Residuals (SSR) represents the total of all squared differences between the observed values and the predicted values :

$$SSR = \sum_{i=1}^{n} \epsilon_i^2 \tag{4}$$

Substituting $\epsilon_i = y_i - \hat{y}_i$, we get:

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

We can also be written as:

$$SSR = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \tag{6}$$

Ordinary least square (OLS)

From the model :

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{7}$$

- y_i : is known
- x_i : is known
- $\epsilon_i = y_i - \alpha - \beta x_i$
- α & β are unknown parameters (intercept & slope)

The purpose of linear regression is to estimate $\hat{\alpha}$ & $\hat{\beta}$ such that the SSR is minimized.

Minimization problem

The minimization problem can be written as :

$$\min_{\alpha,\beta} SSR \tag{8}$$

$$\min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2. \tag{9}$$

There exists two type of solutions. Using iterative algorithm, or finding the analytical solution.

OLS estimator

$$\min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \tag{10}$$

$$\frac{\partial SSR}{\partial \hat{\alpha}} = 0 \tag{11}$$

$$\frac{\partial SSR}{\partial \hat{\beta}} = 0 \tag{12}$$

OLS estimator

By derivating and solving for both parameters, you should find :

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{13}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{14}$$

Assumptions for SLR

In ordinary least squares (OLS) regression, we need assumptions to ensure that the estimated coefficients are **unbiased**, **efficient**, and **consistent**.

1. **Linearity** : The relationship between x and y must be linear.
2. **Independence of errors** : There is not a relationship between the residuals and the y variable; in other words, y is independent of errors.
3. **Normality of errors** : The residuals must be approximately normally distributed.
4. **Homoscedasticity** : The variance of the residuals is the same for all values of x.

Multi-linear regression

So far, we solve the linear equation for the uni-variate case, i.e. we estimated the relation in between one explanatory variable and one target variable.
What if we want to estimate y with more than one x ?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{15}$$

Matrix form

We can redefine the linear regression using matrices :

$$y = X.\beta + \epsilon \tag{16}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} . \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{17}$$

Residuals

As for SLR, we can compute the residuals:

$$y = X\beta + \epsilon \tag{18}$$

$$\epsilon = y - X\hat{\beta} \tag{19}$$

SSR

For the SLR we had :

$$SSR = \sum_{i=1}^{n} \epsilon_i^2 \tag{20}$$

In MLR, the residuals are also in a vector (one residual per observation/prediction). But we can re write it in matrix form as :

$$\epsilon^T . \epsilon = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_n \end{bmatrix} . \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{21}$$

$$\epsilon^T . \epsilon = \begin{bmatrix} \epsilon_1 . \epsilon_1 & \epsilon_2 . \epsilon_2 & \cdots & \epsilon_n . \epsilon_n \end{bmatrix} \tag{22}$$

SSR

Looking at the residuals, and the vector form of SSR, we can develop the SSR :

$$\begin{aligned}\epsilon'\epsilon &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}\end{aligned}$$

Minimization problem for MLR

$$\min_{\hat{\beta}} \sum_{i=1}^{N} \epsilon_i^2 \tag{23}$$

$$\min_{\hat{\beta}} \epsilon' \epsilon \tag{24}$$

$$\min_{\hat{\beta}} y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \tag{25}$$

OLS

By computing the FOC and solving for $\hat{\beta}$, you should find :

$$\hat{\beta} = (X'X)^{-1}X'y \tag{26}$$

Thank you for your attention !