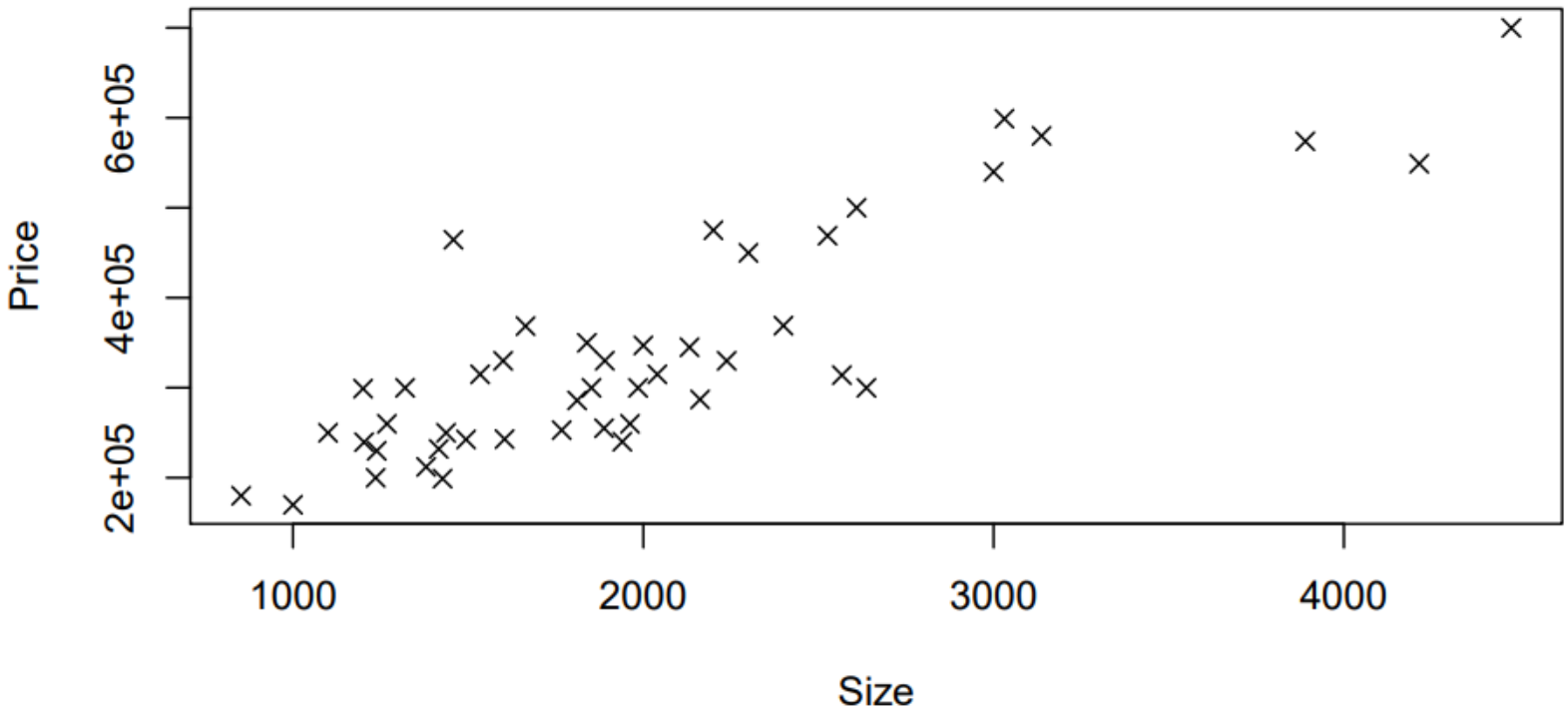

KAZAKH-BRITISH TECHNICAL UNIVERSITY

# Introduction to Machine Learning
# Week 5

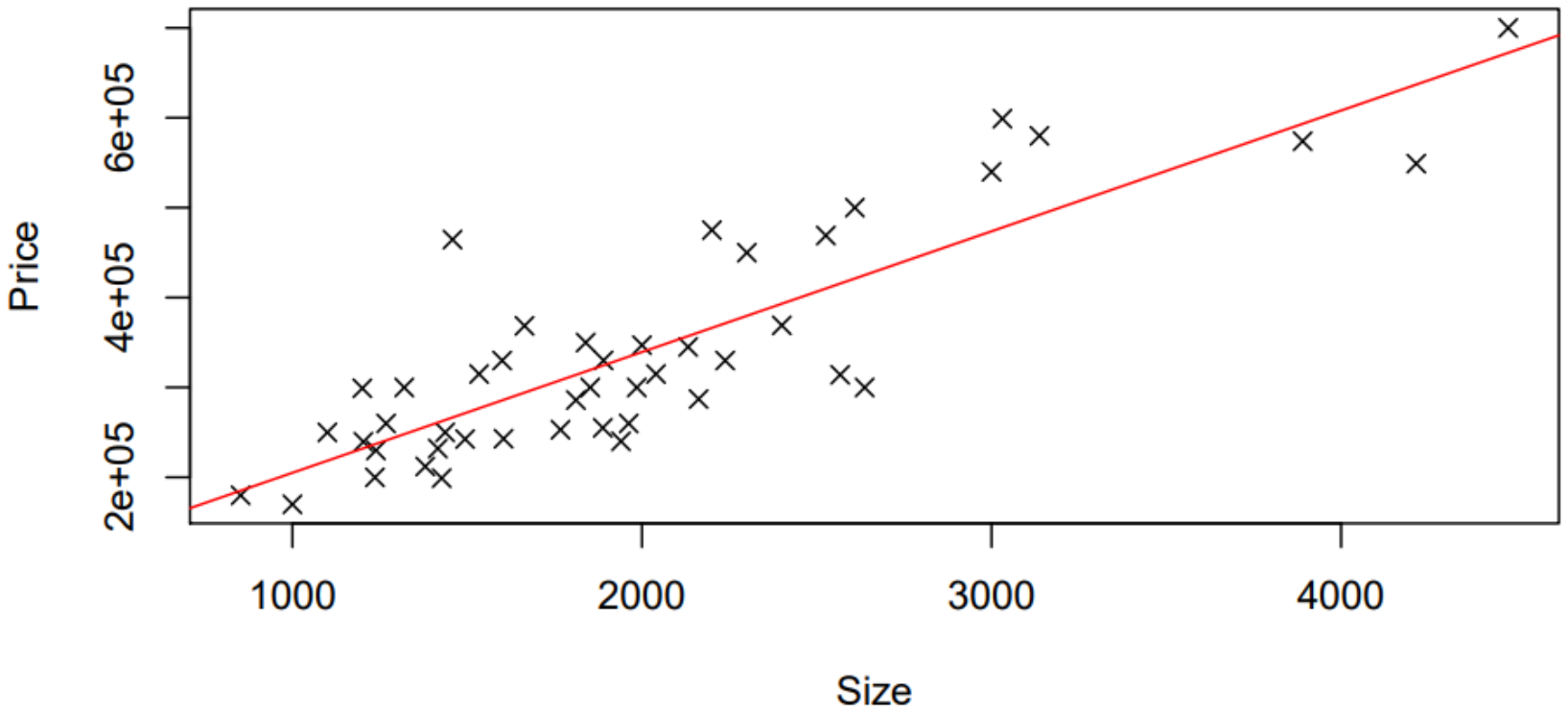
Olivier JAYLET

School of Information Technology and Engineering

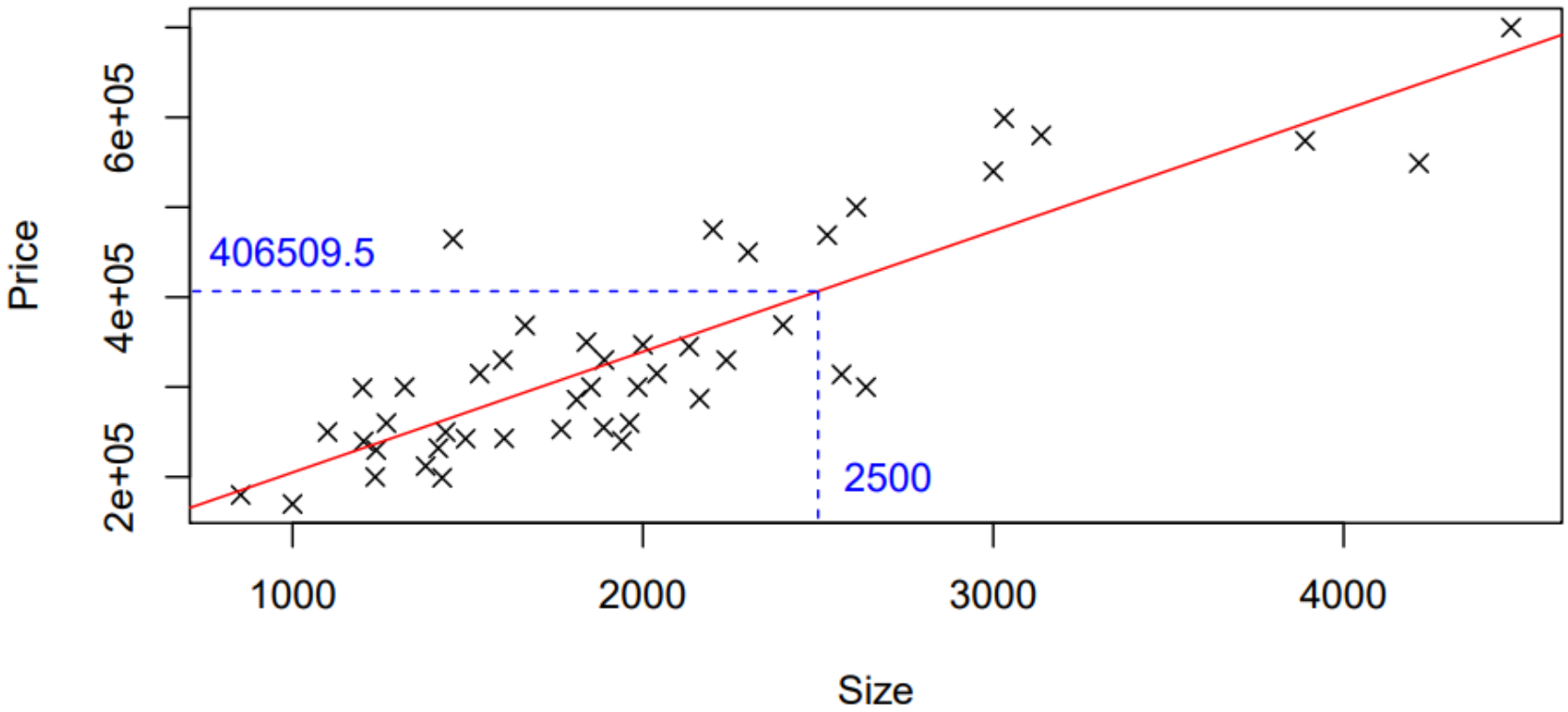
# Context



# 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,
- $\alpha$ is the intercept (constant) of the regression line,
- $\beta$ is the slope of the regression line,
- $\epsilon$ is the error term (or unpredictable random disturbance).

# Representation

# Parameters interpretation

$\alpha$ : Value of the dependent variable ($y$) when the independent variable ($x$) is zero.
$\beta$ : 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 $\epsilon_i$ 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 $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$, 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:

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

It can also be written as:

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