# ML - 6

- Supervised Learning
  - Least-squares based Regression
    - Linear Regression

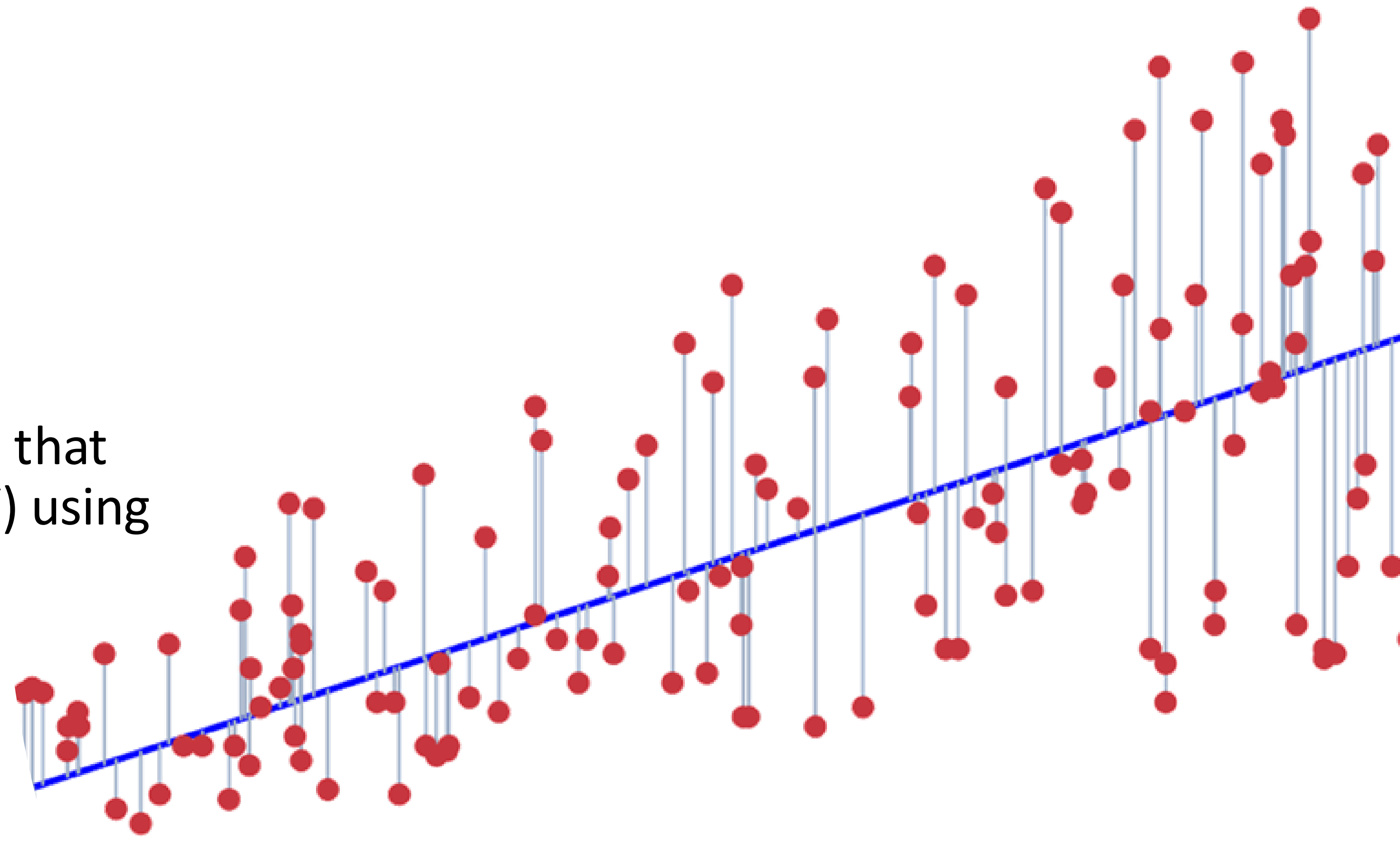
# Machine Learning

Prof. Umut ORHAN

# Linear Regression

Linear regression is a statistical method that aims to predict a dependent variable (Y) using an independent variable (X).

$$Y = W_0 + W_1 X + E$$



$W_0$ *and* $W_1$ represent weights,
$X$ is independent variable,
$E$ is unpredictable error,
$Y$ is dependent variable.

# Linear Regression

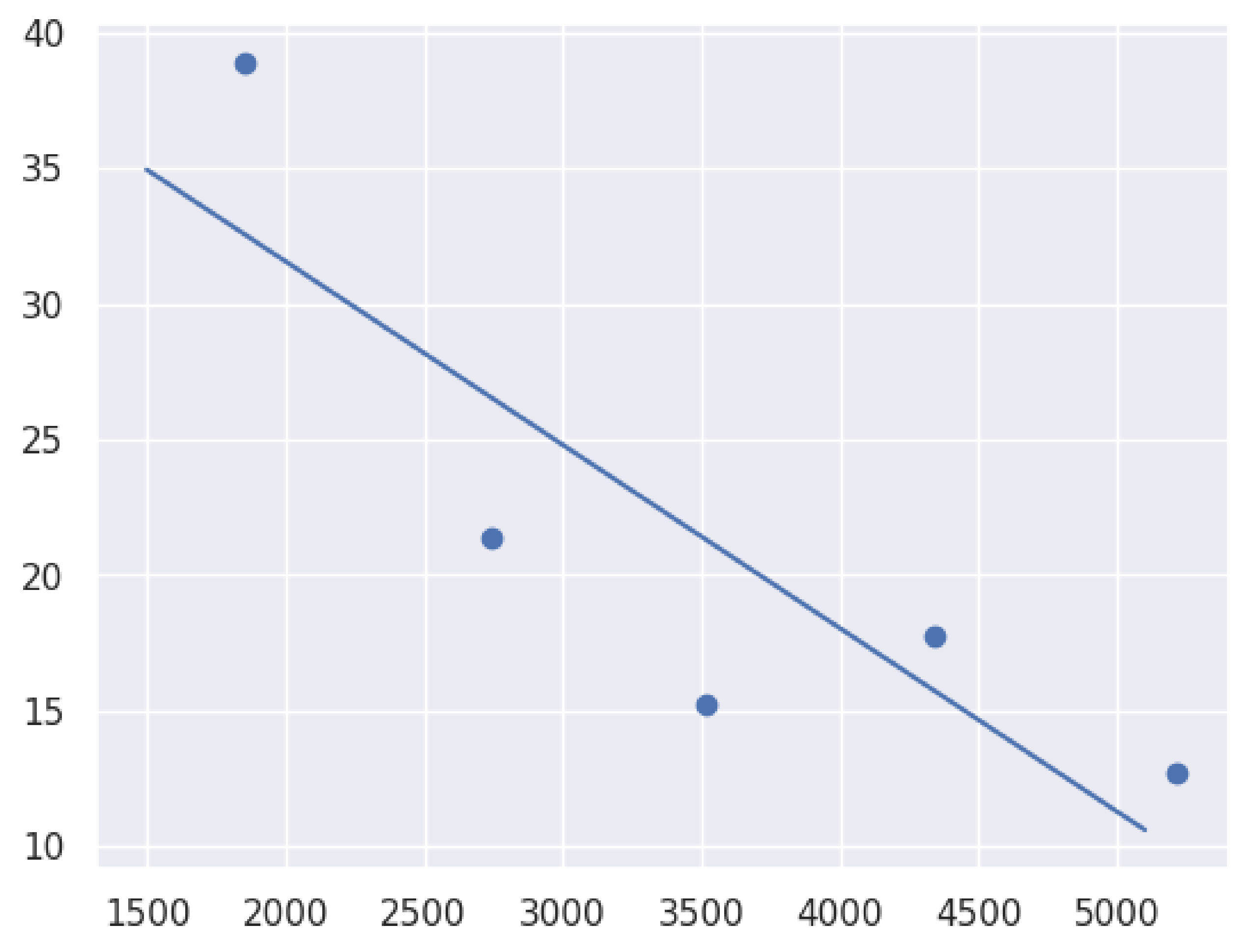
Evaluation of a linear regression model:

$$R^2 = 1 - \frac{\sum(D_i - Y_i)^2}{\sum(D_i - \overline{D})^2}$$

$$MSE = \frac{1}{N}\sum(D_i - Y_i)^2$$

# Sample

| X | D |
|---|---|
| 2743 | 21,4 |
| 3518 | 15,2 |
| 1855 | 38,9 |
| 5214 | 12,7 |
| 4341 | 17,8 |




$$Y = 45.11 - 0.0067\ X$$

$R^2=0.731$ and $MSE=23.30$

# Multiple Linear Regression

Real world problems often focus on representing a single dependent variable with multiple independent variables.

$$Y = W_0 + \sum_{j=1}^{M} W_j X_j + E$$

# Ordinary Least Squares Method

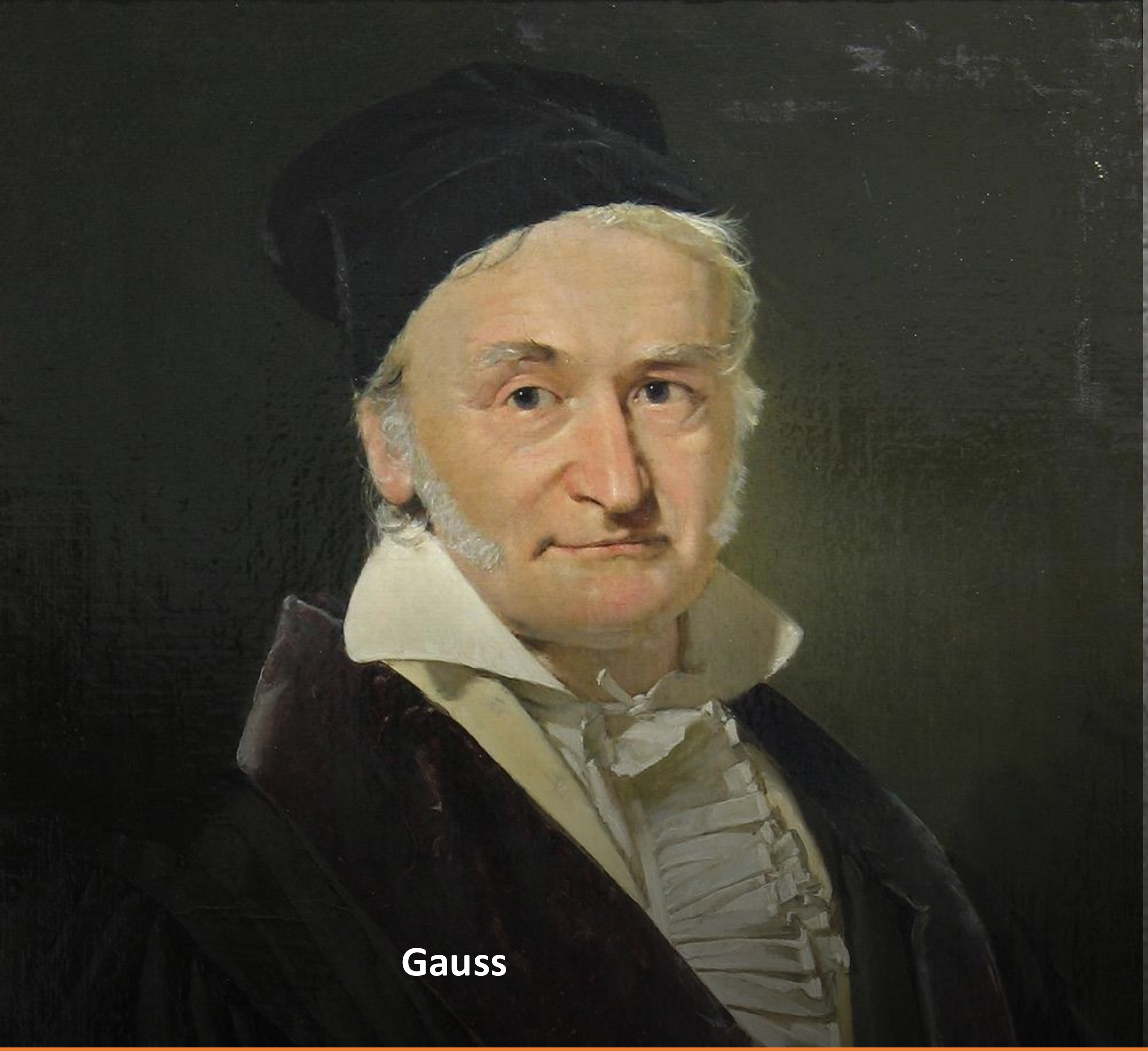
The goal of OLS is to find the weights in a multiple linear regression model that will minimize the sum of squares of the error terms.

$$min \sum_{i=1}^{N} e^2 = min \sum_{i=1}^{N} \left( d_i - \sum_{j=1}^{M} w_j x_{ij} \right)^2$$

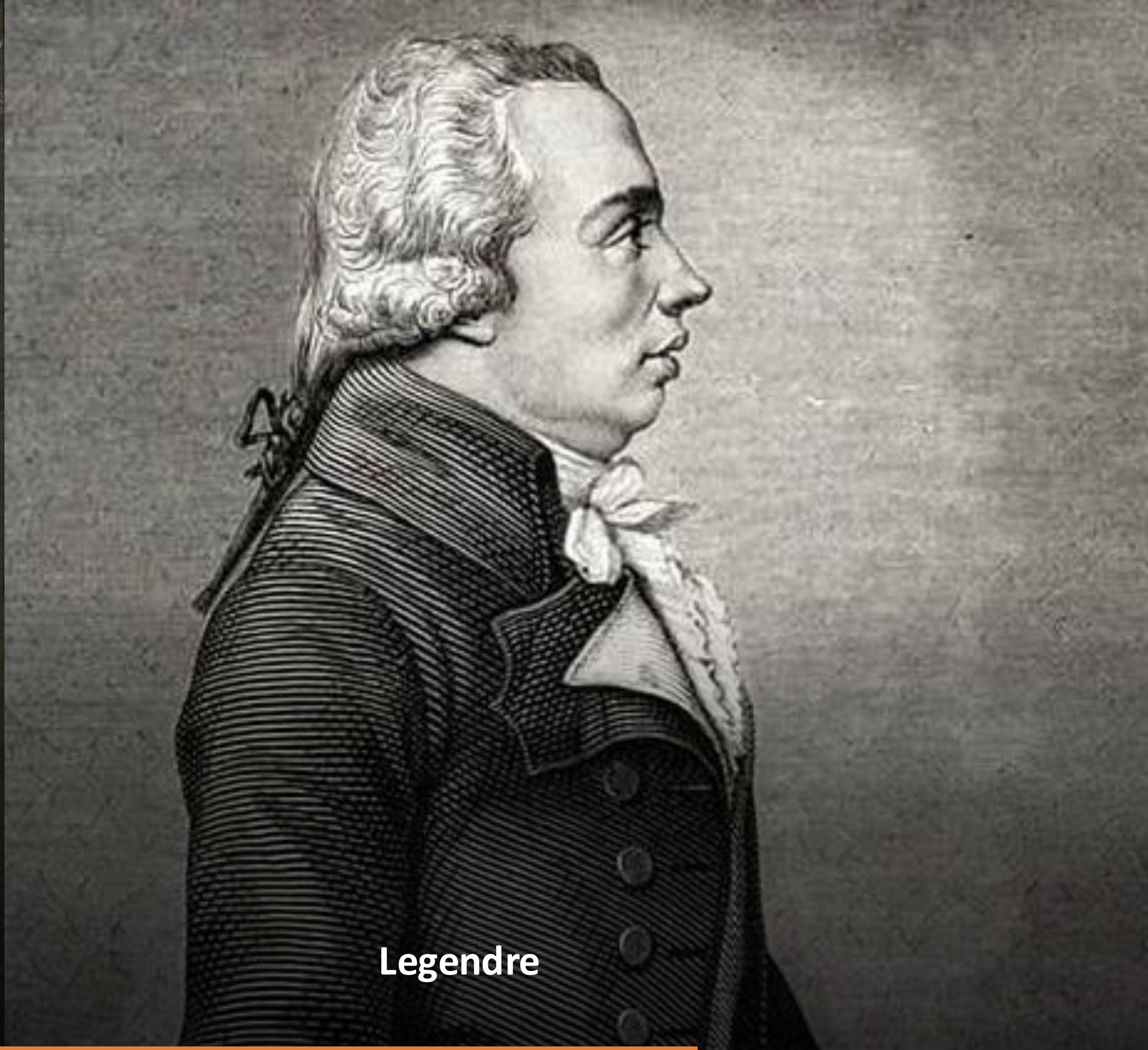
# Ordinary Least Squares Method

The equation that allows us to find the weights that obtain the least squared errors is as follows.

$$\mathbf{W} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{D}$$



Gauss



Legendre

# Who is the Inventor of OLS?

# Using Regression for Classification

For a two-class dataset, if the class labels are chosen as 0 and 1, using a threshold such as 0.5 is sufficient to convert a continuous value into discrete classes.

If the calculated value is greater than the threshold, 1, otherwise 0.

# Sample

The IRIS dataset has 4 features, also called independent variables, and a dependent variable with 3 classes.

We map dependent variables with more than two classes to two-class labels with a "one class vs all others" perspective.

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | D |
|---|---|---|---|---|
| 5.1 | 3.5 | 1.4 | 0.2 | A |
| 4.9 | 3 | 1.4 | 0.2 | A |
| 7 | 3.2 | 4.7 | 1.4 | B |
| 6.4 | 3.2 | 4.5 | 1.5 | B |
| 6.3 | 3.3 | 6 | 2.5 | C |
| 5.8 | 2.7 | 5.1 | 1.9 | C |

# Sample

The "one class vs all others" perspective requires us to transform the 3-class variable D into three different 2-class variables.

In the first equation labeled D1, we will try to solve the "A class vs others" problem.

| D | $D_1$ | $D_2$ | $D_3$ |
|---|---|---|---|
| A | 1 | 0 | 0 |
| A | 1 | 0 | 0 |
| B | 0 | 1 | 0 |
| B | 0 | 1 | 0 |
| C | 0 | 0 | 1 |
| C | 0 | 0 | 1 |

# Sample

Now, we have three multiple linear regression equations:

- $D_1 = 1.2 + 0.29X_1 - 0.16X_2 - 1.02X_3 + 1.43X_4$
- $D_2 = -3.4 + 0.82X_1 - 0.2X_2 + 0.04X_3 - 0.52X_4$
- $D_3 = 3.2 - 1.11X_1 + 0.35X_2 + 0.98X_3 - 0.92X_4$

Computed mean squared errors (MSEs);

- for $D_1$ , MSE=0.030,
- for $D_2$ , MSE=0.301, (this is the worst one)
- for $D_3$ , MSE=0.014

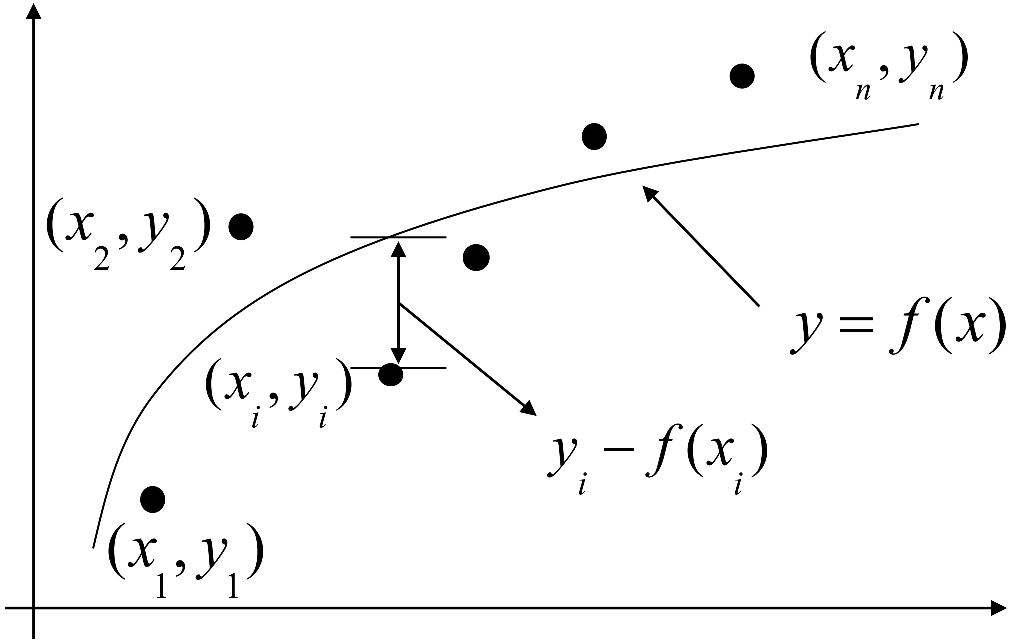
# Sample

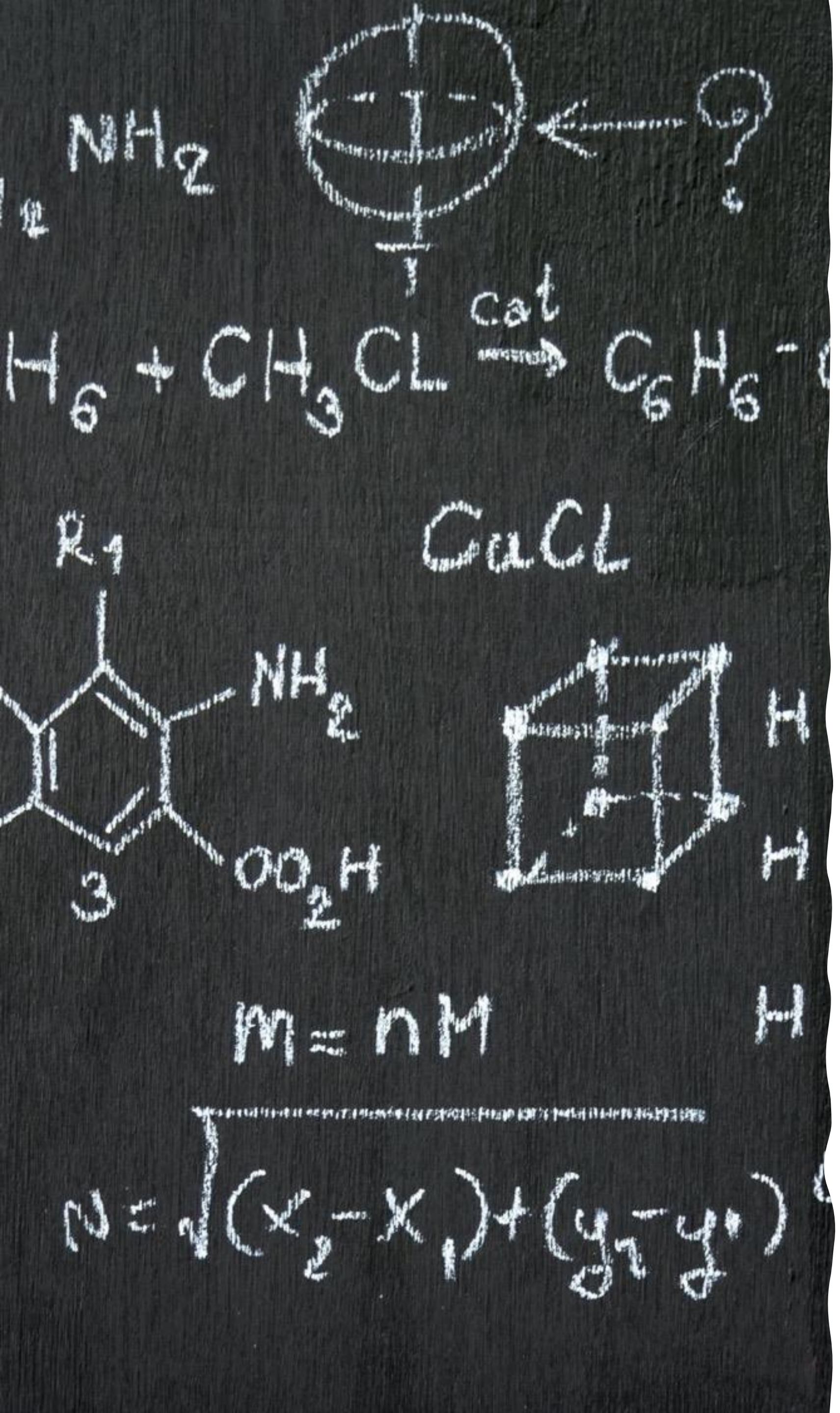
When the output values are calculated for each data sample using the linear regression equations with determined weights, the Y values in the table are calculated. In each row, the class representation that wins is determined with "the biggest wins" principle.

| D | $Y_1$ | $Y_2$ | $Y_3$ |
|---|---|---|---|
| A | 0.98 | 0.06 | -0.04 |
| A | 1.00 | -0.01 | 0 |
| B | -0.06 | 1.19 | -0.13 |
| B | 0.11 | 0.63 | 0.25 |
| C | -0.02 | 0.07 | 0.95 |
| C | -0.02 | 0.05 | 0.96 |


# Nonlinear Regression

Nonlinear regression is used when the relationship between the dependent and independent variables is not linear.






# Nonlinear Regression

There are many kind of nonlinear model. Some of them are:

1. Exponential model: $y = ae^{bx}$
2. Power model: $y = ax^b$
3. Saturation growth model: $y = \frac{ax}{b+x}$
4. Polynomial model: $y = a_0 + a_1 x + \cdots + a_m x^m$

# Manual Calculations

By using the OLS method, find the coefficients of equation.

| $X_1$ | $X_2$ | D |
| --- | --- | --- |
| 2 | 3 | 1 |
| 5 | 4 | 3 |
| 1 | 5 | 4 |

# Manual Calculations

1

$$X = \begin{bmatrix} 2 & 3 \\ 5 & 4 \\ 1 & 5 \end{bmatrix} \quad D = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

$$W = (X^T X)^{-1} X^T D$$

$$\begin{bmatrix} 2 & 5 & 1 \\ 3 & 4 & 5 \end{bmatrix} * \begin{bmatrix} 2 & 3 \\ 5 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 30 & 31 \\ 31 & 50 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 5 & 1 \\ 3 & 4 & 5 \end{bmatrix} * \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 21 \\ 35 \end{bmatrix}$$

2

$$\begin{bmatrix} 30 & 31 & 21 \\ 31 & 50 & 35 \end{bmatrix} = \begin{bmatrix} 30 & 31 & 21 \\ 1 & 19 & 14 \end{bmatrix} = \begin{bmatrix} 0 & 539 & 399 \\ 1 & 19 & 14 \end{bmatrix} = \begin{bmatrix} 0 & 1 & \frac{399}{539} \\ 1 & 19 & 14 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 19 & 14 \\ 0 & 1 & 399/539 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 14 - 19x\,\frac{399}{539} \\ 0 & 1 & 399/539 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -0.065 \\ 0 & 1 & 0.74 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & w_1 \\ 0 & 1 & w_2 \end{bmatrix}$$

$Y=-0.065*X_1+0.74*X_2$

3

In order find MSE,

$$\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 & 3 \\ 5 & 4 \\ 1 & 5 \end{bmatrix} * \begin{bmatrix} -0.065 \\ 0.74 \end{bmatrix} = \begin{bmatrix} -1.1 \\ 0.36 \\ 0.36 \end{bmatrix}$$

MSE=0.48

# Machine Learning

## 6. week

Thanks for watching