**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI – HYDERABAD CAMPUS**
**BITS F464 : MACHINE LEARNING**
**FIRST SEMESTER 2023-2024**

# A worked-out example on Polynomial Regression

**Here's a step-by-step solved numerical problem on polynomial regression of degree 2.**

Let's assume the following dataset related to the relationship between the temperature (in Celsius) and the corresponding pressure (in kPa) of a gas at a certain altitude:

| Temperature (Celsius) | Pressure (kPa) |
|---|---|
| 10 | 95 |
| 20 | 89 |
| 30 | 75 |
| 40 | 60 |
| 50 | 50 |

We want to perform polynomial regression of degree 2 to find the best fit curve that represents the relationship between temperature and pressure. The polynomial regression model of degree 2 is given by the equation:

$$y = ax^2 + bx + c$$

where:

- $y$ is the dependent variable (pressure in this case)
- $x$ is the independent variable (temperature in this case)
- $a, b, c$ are the coefficients to be determined.

Step 1: Formulate the polynomial regression problem using matrices:

Let's define the following matrices:

$$X = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ x_4^2 & x_4 & 1 \\ x_5^2 & x_5 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

$$\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

where:

- $x_i$ is the temperature value for the $i$-th data point
- $y_i$ is the pressure value for the $i$-th data point.

Step 2: Calculate the values for $X$, $Y$, and $\theta$:

From the dataset, we have:

$$x_1 = 10,\ x_2 = 20,\ x_3 = 30,\ x_4 = 40,\ x_5 = 50$$

$$y_1 = 95,\ y_2 = 89,\ y_3 = 75,\ y_4 = 60,\ y_5 = 50$$

So, the matrices become:

$$X = \begin{bmatrix} 100 & 10 & 1 \\ 400 & 20 & 1 \\ 900 & 30 & 1 \\ 1600 & 40 & 1 \\ 2500 & 50 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} 95 \\ 89 \\ 75 \\ 60 \\ 50 \end{bmatrix}$$

$$\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Step 3: Calculate the optimal values for $\theta$ using the normal equation:

The normal equation for polynomial regression is given by:

$$\theta = (X^T X)^{-1} X^T Y$$

where:

- $X^T$ is the transpose of matrix $X$
- $(X^T X)^{-1}$ is the inverse of the matrix $(X^T X)$

Let's calculate $(X^T X)$:

$$X^T X = \begin{bmatrix} 9790000 & 225000 & 5500 \\ 225000 & 5500 & 150 \\ 5500 & 150 & 5 \end{bmatrix}$$

Next, calculate $(X^T X)^{-1}$:

$$(X^TX)^{-1} = \begin{bmatrix} 9790000 & 225000 & 5500 \\ 225000 & 5500 & 150 \\ 5500 & 150 & 5 \end{bmatrix}^{-1}$$

You can use a matrix calculator or software to find the inverse. For this example, we have:

$$(X^TX)^{-1} = \begin{bmatrix} 0.000007 & -0.000429 & 0.005 \\ -0.000429 & 0.026714 & -0.33 \\ 0.005 & -0.33 & 4.6 \end{bmatrix}$$

Now, calculate $\theta = (X^TX)^{-1}X^TY$. After performing the matrix multiplication, you will get the values for $\theta$:

$$\theta = \begin{bmatrix} -0.05832 \\ -0.95008 \\ 105 \end{bmatrix}$$

So, the polynomial regression model of degree 2 is approximately:

$$y = -0.05832x^2 - 0.95008x + 105$$

This model represents the best-fit curve that fits the dataset in a polynomial fashion of degree 2. You can now use this equation to predict pressure values for different temperature values.