Polynomial Regression – Simplified Theory

1. Introduction

- **Polynomial Regression** is a type of regression that **extends Linear Regression** by adding **non-linear relationships** between the independent variable(s) and the dependent variable.
- It allows a model to **fit curves** to the data instead of just a straight line.

2. What is Polynomial Regression?

- Standard **Linear Regression equation**:

$$y = \beta_0 + \beta_1 x + \epsilon$$

This assumes a **straight-line relationship** between x and y.

- In **Polynomial Regression**, we include **higher-degree terms** of x:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \epsilon$$

Where:

- $x^2, x^3, \ldots x^n$ = **polynomial terms** that capture **curvature** in data.
- $\beta_0, \beta_1, \ldots, \beta_n$ = model coefficients.

3. Why Polynomial Regression?

- Many real-world relationships are **not perfectly linear**.
- **Example:**
    1. Predicting **growth of a plant** with time may follow a **curved pattern**.
    2. Predicting **sales revenue** based on advertising spend may have **diminishing returns**.
- Linear regression cannot capture these patterns, but polynomial regression can.

4. How It Works

- The data is **transformed** by creating **additional features**:
  - From a single variable x, we create:
    - $x^2, x^3, \ldots, x^n$
- We then apply **Linear Regression** on these new features:

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$$

Though the relationship is non-linear, the model remains **linear in coefficients**, making it solvable using linear regression techniques.

---

5. Degree of Polynomial

- **Degree 1:** Linear regression (straight line).
- **Degree 2:** Quadratic regression (parabolic curve).
- **Degree 3+:** More complex curves.

Choosing the degree:

- **Too low:** Underfitting (model too simple).
- **Too high:** Overfitting (model too complex, memorizing noise).

---

6. Advantages

- Captures **non-linear patterns** in data.
- Simple to implement using linear regression techniques.
- Works well when **relationship between variables is curved**.

---

7. Disadvantages

- Higher degrees can lead to:
  1. **Overfitting**, especially with small datasets.
  2. Poor performance on unseen data.
- **Extrapolation risk:** Predictions outside data range can behave unpredictably.
- **Multicollinearity:** Polynomial terms can be highly correlated.

➢ Key Takeaway

- Polynomial regression:
  1. **Extends linear regression** by adding powers of features.
  2. Allows fitting **non-linear relationships** while keeping a **linear model structure**.
- **Formula:**

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$$