POLYNOMIAL REGRESSION

- Polynomial Regression is a form of linear regression where the relationship between the independent variable (x) and the dependent variable (y) is modelled as an nth degree polynomial.
- It is useful when the data exhibits a non-linear relationship allowing the model to fit a curve to the data.

Need for Polynomial Regression

- Non-linear Relationships: Polynomial regression is used when the relationship between the independent variable (input) and dependent variable (output) is non-linear. Unlike linear regression which fits a straight line, it fits a polynomial equation to capture the curve in the data.
- Better Fit for Curved Data: When a researcher hypothesizes a curvilinear relationship, polynomial terms are added to the model. A linear model often results in residuals with noticeable patterns which shows a poor fit. It can capture these non-linear patterns effectively.
- Flexibility and Complexity: It does not assume all independent variables are independent. By introducing higher-degree terms, it allows for more flexibility and can model more complex, curvilinear relationships between variables.

How does a Polynomial Regression work?

- Polynomial regression is an extension of linear regression where higher-degree terms are added to model non-linear relationships.
- The general form of the equation for a polynomial regression of degree n is:

- where:
- y is the dependent variable.
- x is the independent variable.
- β0,β1,…,βn are the coefficients of the polynomial terms.
- n is the degree of the polynomial.
- ϵ represents the error term.
-

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

- The goal of regression analysis is to model the expected value of a dependent variable y in terms of an independent variable x.
- In simple linear regression, this relationship is modeled as:

- y=a+bx+e
- Here
- y is a dependent variable
- a is the y-intercept,
- b is the slope
- e is the error term
-

- However in cases where the relationship between the variables is nonlinear
- (e.g) modelling chemical synthesis based on temperature, a linear model may not be sufficient
- we use polynomial regression which introduces higher-degree terms such as x2 to better capture the relationship.
- For example, a quadratic model can be written as:
    - y=a+b1 x+b2x2+e
    - Here:
    - y is the dependent variable on x
    - a is the y-intercept and
    - e is the error rate.

- In general, polynomial regression can be extended to the nth degree:

$$y = a + b_1x + b_2x^2 + ... + b_nx^n$$

Types of Polynomials

Linear —————— $ax + b = 0$

Quadratic ———— $ax^2 + bx + c = 0$

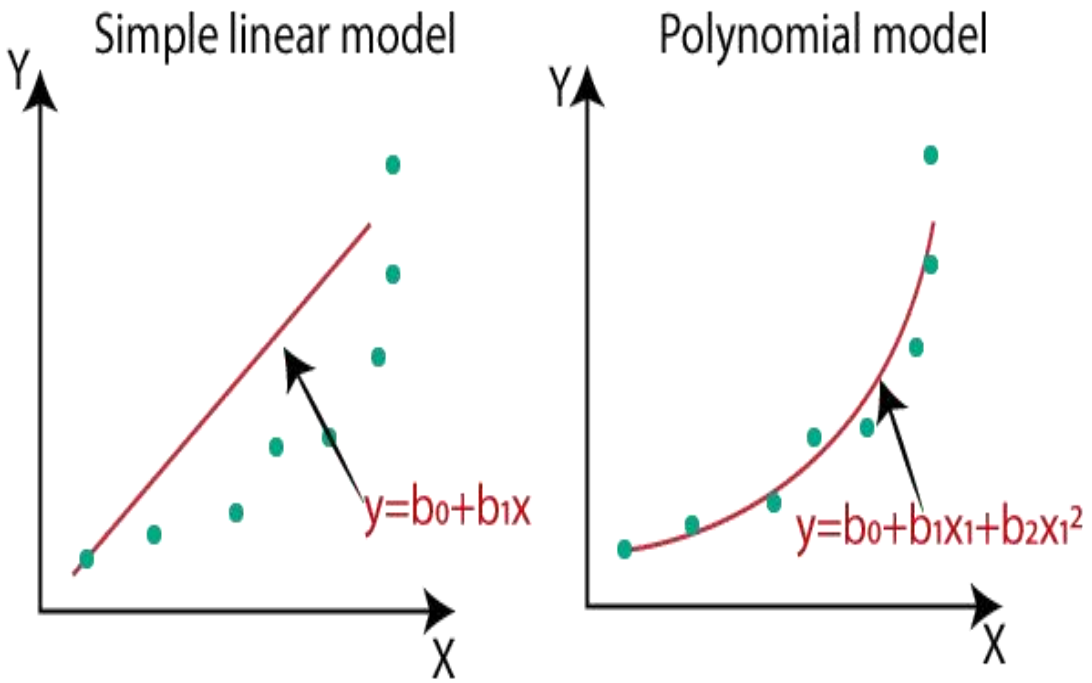
Cubic —— $ax^3 + bx^2 + cx + d = 0$

Assumptions of Polynomial Regression:

- The behavior of a dependent variable can be explained by a linear, or curvilinear, additive relationship between the dependent variable and a set of k independent variables (xi, i=1 to k).
- The relationship between the dependent variable and any independent variable is linear or curvilinear (specifically polynomial).
- The independent variables are independent of each other.
- The errors are independent, normally distributed with mean zero and a constant variance (OLS).

Why do we need Polynomial Regression?

- Let's consider a case of Simple Linear Regression.
- We make our model and find out that it performs very badly,
- We observe between the actual value and the best fit line,which we predicted and it seems that the actual value has some kind of curve in the graph and our line is no where near to cutting the mean of the points.
- This where polynomial Regression comes to the play,it predicts the best fit line that follows the pattern(curve) of the data,as shown in the pic below:



- Polynomial Regression does not require the relationship between the independent and dependent variables to be linear in the data set, This is also one of the main difference between the Linear and Polynomial Regression.

- Polynomial Regression is generally used when the points in the data are not captured by the Linear Regression Model and the Linear Regression fails in describing the best result clearly.

- As we increase the degree in the model,it tends to increase the performance of the model.
- However,increasing the degrees of the model also increases the risk of over-fitting and under-fitting the data.

How to find the right degree of the equation?

- In order to find the right degree for the model to prevent over-fitting or under-fitting, we can use:
-
- Forward Selection:
- This method increases the degree until it is significant enough to define the best possible model.
- Backward Selection:
- This method decreases the degree until it is significant enough to define the best possible model.

Cost Function of Polynomial Regression

- Cost Function is a function that measures the performance of a Machine Learning model for given data.
- Cost Function is basically the calculation of the error between predicted values and expected values and presents it in the form of a single real number.
- Cost Function of Polynomial Regression can also be taken to be Mean Squared Error,However there will a slight change in the equation.

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

Example

- Using the following dataset, fit a second-degree polynomial regression model

Y=a+bX+cX2

- Compute the required summations.
- Form the normal equations.

- Determine the regression equation.
- Predict the value of Y when X=6.

X	1	2	3	4	5
Y	3	5	10	18	29

Step 1: Prepare the Calculation Table

- Also,
- n=5

X	Y	X^2	X^3	X^4	XY	X^2Y
1	3	1	1	1	3	3
2	5	4	8	16	10	20
3	10	9	27	81	30	90
4	18	16	64	256	72	288
5	29	25	125	625	145	725
∑X=15	65	55	225	979	260	1126

Step 2: Form the Normal Equations

- The quadratic model is:
  - Y=a+bX+cX2

The normal equations are:

5a+15b+55c=65
15a+55b+225c=260
55a+225b+979c=1126

Step 3: Solve the Equations

Solving the three equations gives:
- a=5
- b=−3
- c=2

Therefore, the fitted polynomial regression equation is:

Y=5−3X+2X2

Step 4: Predict for X=6

- Substitute X=6:
- Y=5−3(6)+2(6)2
- Y=5−18+72

- Y=59

Final Answer

Polynomial Regression Equation
- Y=5−3X+2X2

Predicted value when X=6
- Y=59

Implementation in Python

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures

# Sample data
X = np.array([1, 2, 3, 4, 5]).reshape(-1, 1)
y = np.array([2, 6, 14, 28, 45])

# Polynomial transformation (degree=2)
poly = PolynomialFeatures(degree=2)
X_poly = poly.fit_transform(X)

# Fit model
model = LinearRegression()
model.fit(X_poly, y)

# Predict
y_pred = model.predict(X_poly)
plt.scatter(X, y, color='blue')
plt.plot(X, y_pred, color='red')
plt.show()
```