Machine Learning Diploma

Level3: Machine Learning

Session 2

Agenda

- ➔ Linear Regression with Multiple variables
- ➔ Polynomial Regression
- ➔ Implementation of Polynomial Regression
- ➔ Implementation of Multivariate Linear Regression
- ➔ Sklearn Linear Regression

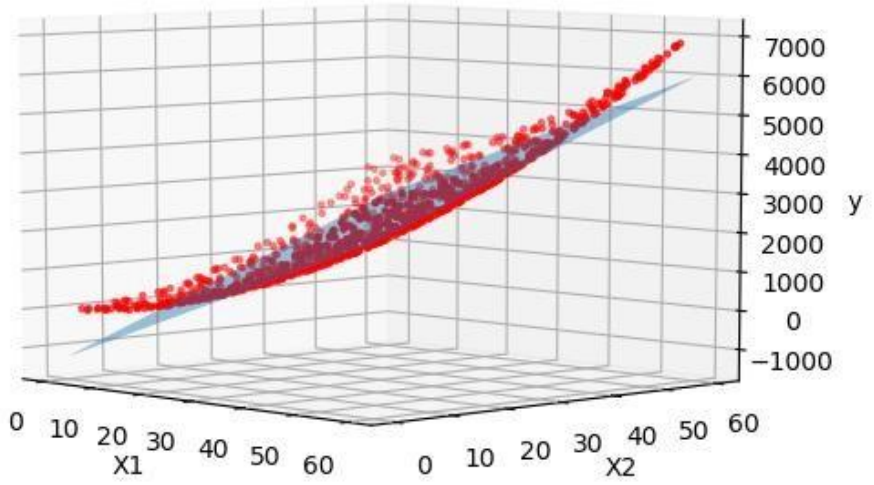
1. Linear Regression with Multiple Variables

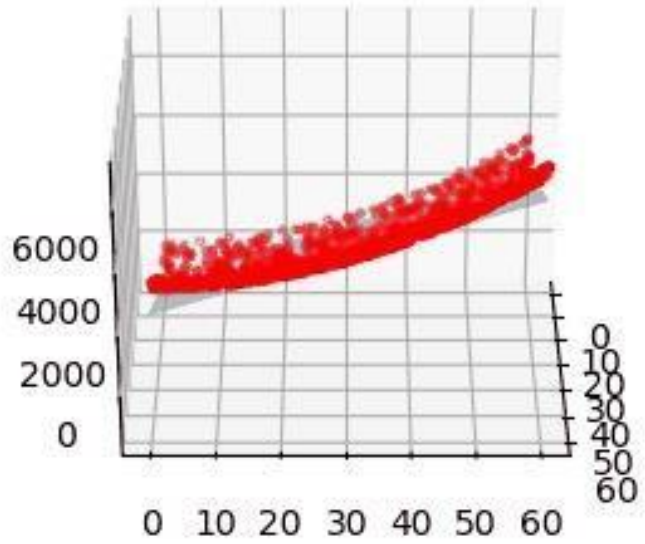
AMIT

Linear Regression with Multiple Variables:

➔ Multivariate Linear Regression uses several explanatory (independent) variables to predict the outcome of a response (dependent) variable.

➔ $y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots\ldots a_n x_n$

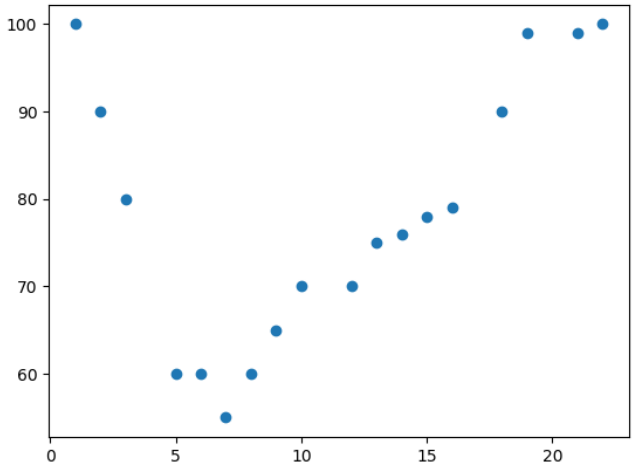


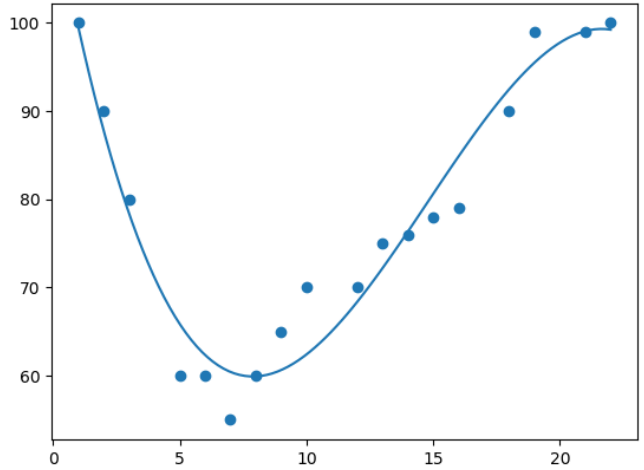


2. Polynomial Regression

Polynomial Regression:

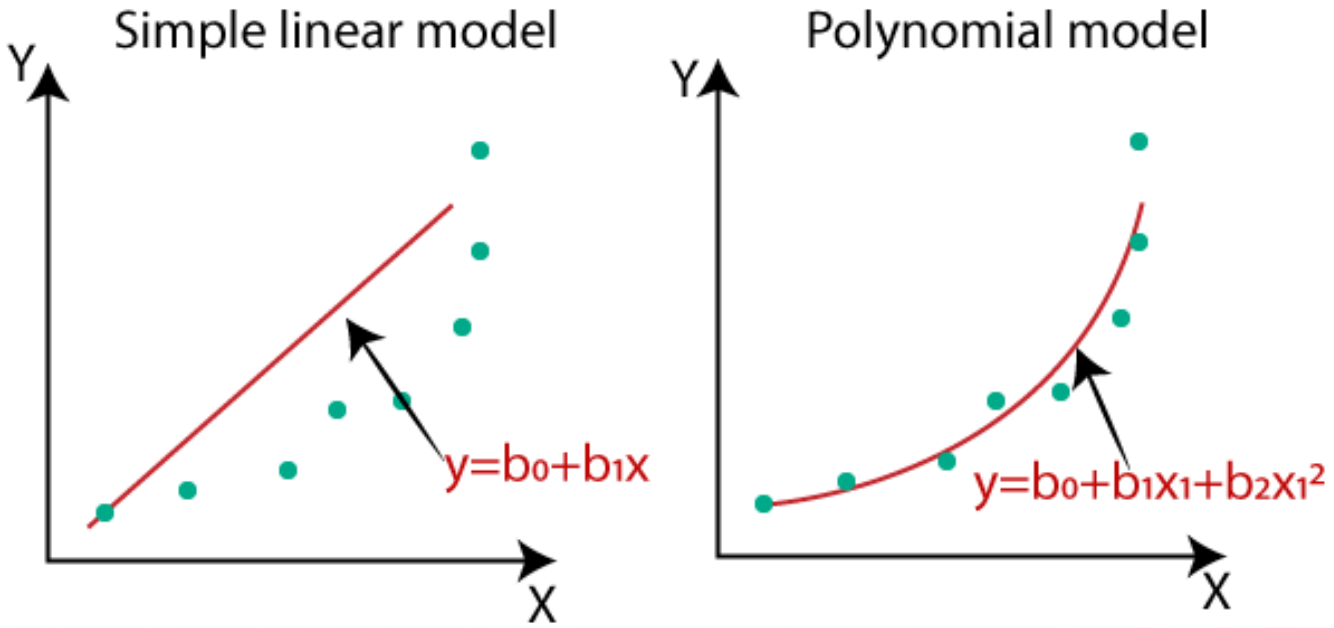
- ➔ If your data points clearly will not fit a linear regression (a straight line through all data points), it might be ideal for polynomial regression.
- ➔ Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial.





Polynomial Regression:

- ➔ It is a linear model with some modification in order to increase the accuracy.
- ➔ The dataset used in Polynomial regression for training is of non-linear nature.

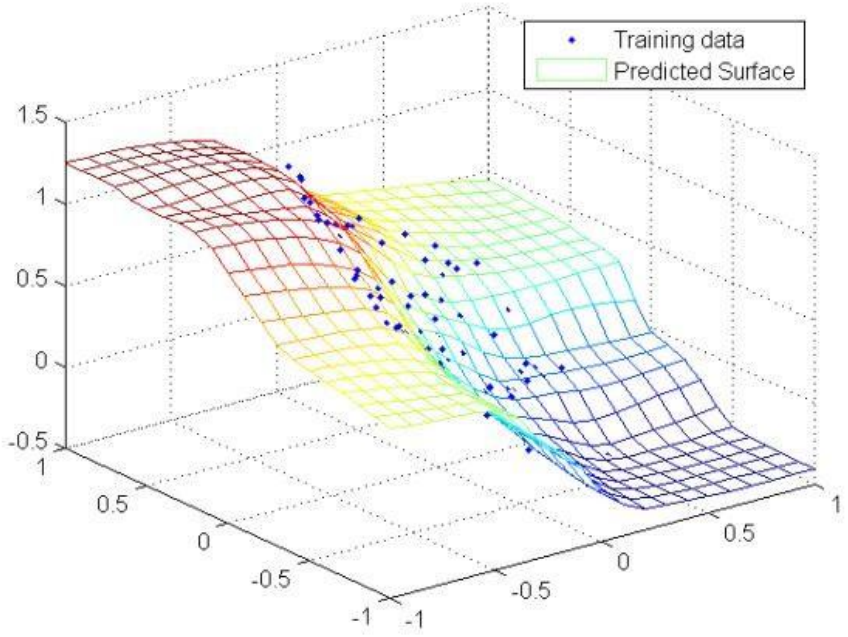


AMIT

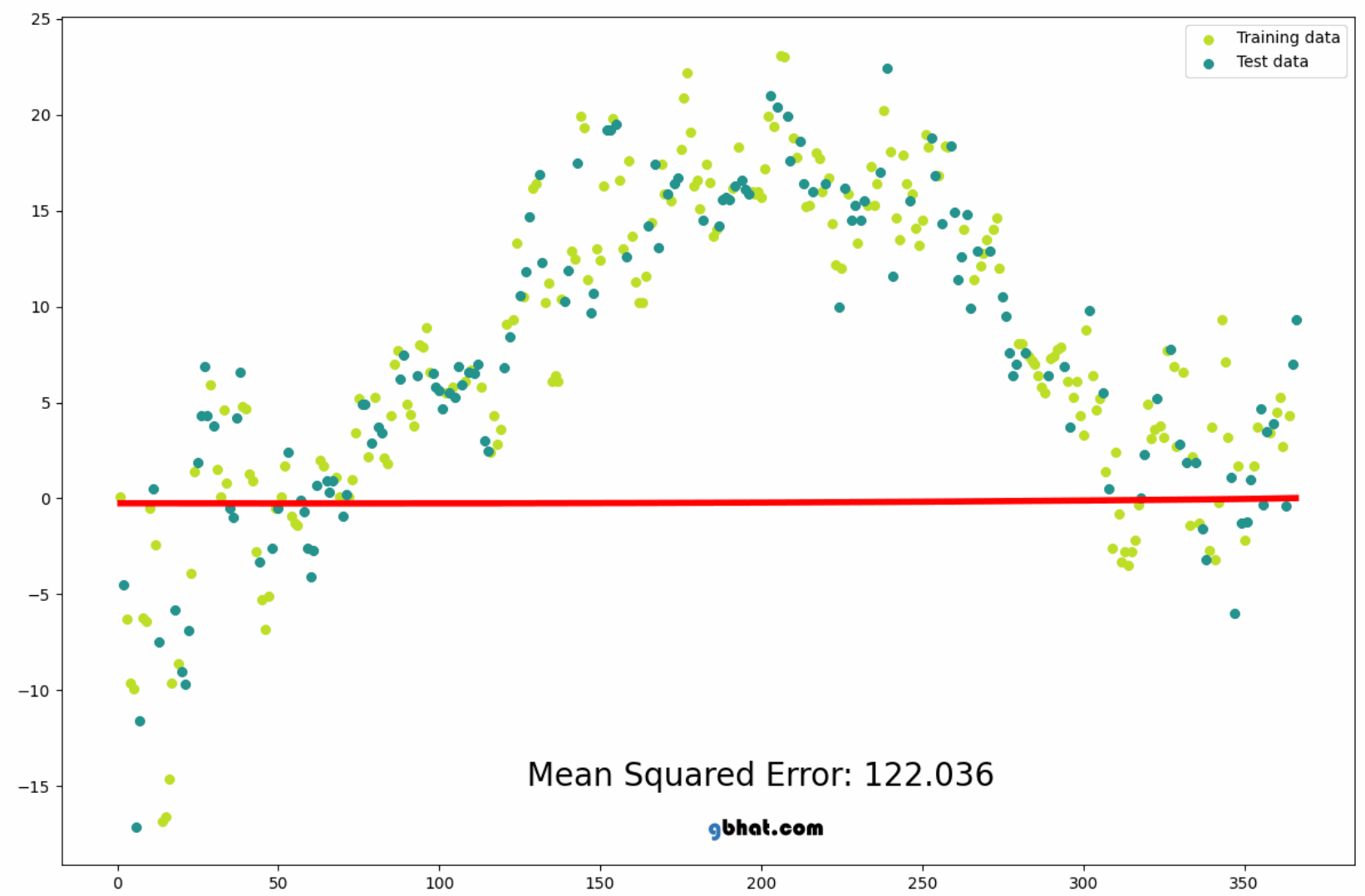
Polynomial Regression:

➔ The original features are converted into polynomial features of required degree (2,3,..,n) and then modelled using a linear model.

➔ $y = a_0 + a_1x_1 + a_2x_1^2 + a_3x_1^3 + \ldots\ldots a_nx_1^n + a_{n+1}x_2 + a_{n+2}x_2^2 + a_{n+3}x_2^3 + \ldots\ldots a_{2n}x_2^n$



Polynomial Regression:



3. Implementation of Linear Regression

Null Hypothesis:

- A **null hypothesis** is a hypothesis that says there is no statistical significance between the two variables in the hypothesis.
- **Null hypothesis** would be something like this: There is no statistically significant relationship between the type of water I feed the flowers and growth of the flowers.
- The **null hypothesis** is nearly always "something didn't happen" or "there is no effect" or "there is no relationship".
- The usual method is to test the null at some significance level (most often, **0.05**)

$$Y = a_0 + \cancel{a_1 X}$$

AMIT

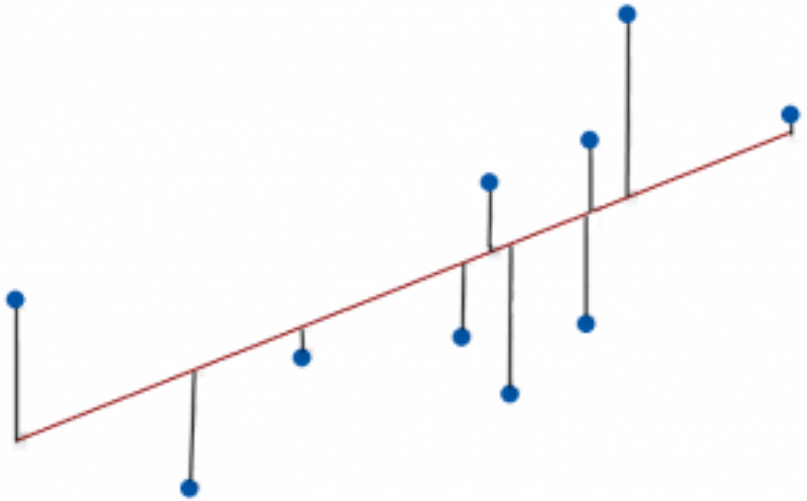
P Value Significance Level:

Significance Level		Specification
$p > 0.05$		not significant
$p \leq 0.05$	(5%)	significant
$p \leq 0.01$	(1%)	very significant
$p \leq 0.001$	(0.1%)	highly significant

AMIT

Linear Regression Output Result:

- **slope**: Slope of the regression line (a_o).
- **intercept**: Intercept of the regression line (a_1).
- **P_value**: The p-value for a hypothesis test whose null hypothesis is that the slope is zero.
- **R_value**: Correlation coefficient that measures the strength of the linear relationship .
- **stderr**: Standard error of the estimated slope (gradient) represents the average distance that your observed values deviate from the regression line.

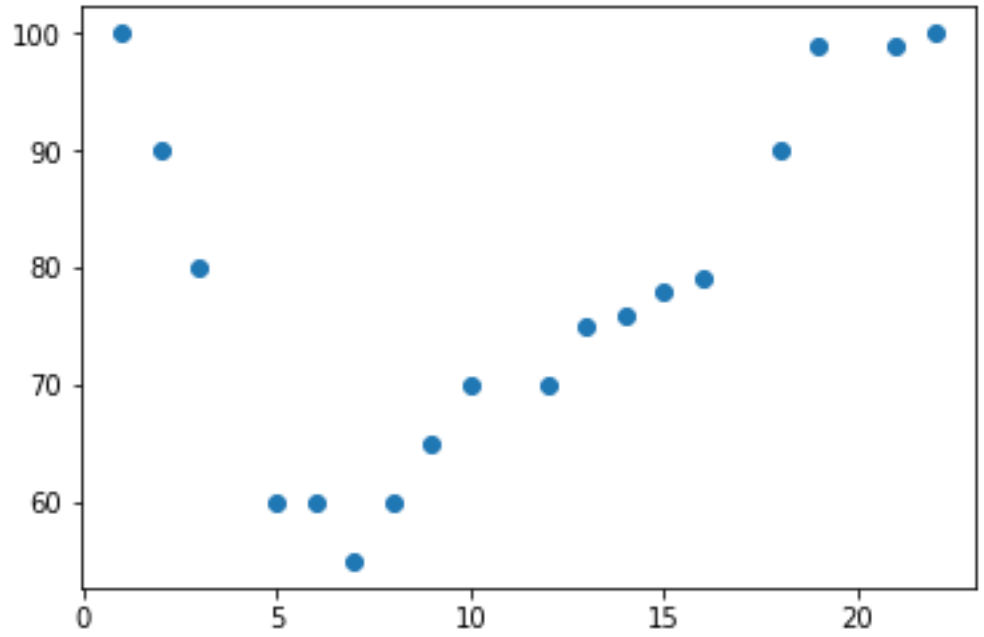


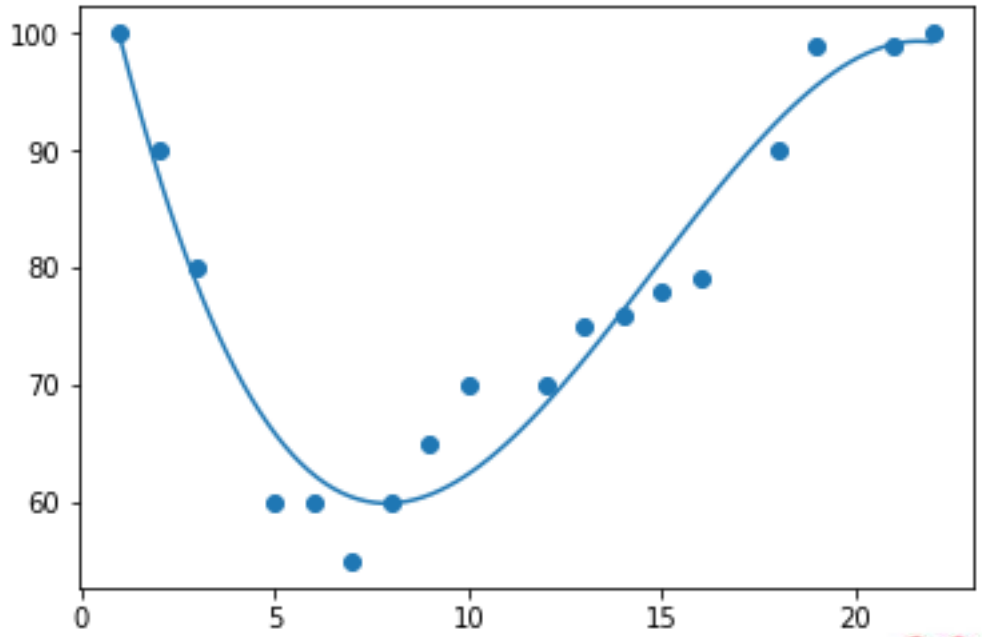
AMIT

4. Implementation of Polynomial Regression

Polynomial Regression Output Result:

- ➔ **polyfit:** Fit x & y to get Polynomial Coefficients for nth order.
- ➔ **poly1d:** Construct the Polynomial Equation





5. Sklearn Linear Regression

5. Implementation of Multivariate Linear Regression

Multivariate Linear Regression Example:

Car	Model	Volume	Weight	CO2
Toyota	Aygo	1000	790	99
Mitsubishi	Space Star	1200	1160	95
Skoda	Citigo	1000	929	95
Fiat	500	900	865	90
Mini	Cooper	1500	1140	105
VW	Up!	1000	929	105
Skoda	Fabia	1400	1109	90
Mercedes	A-Class	1500	1365	92

Any Questions?

AMIT

THANK YOU!

AMIT