When to use regression?

If target variable is a continuous numeric variable (100–2000), then use a regression algorithm.

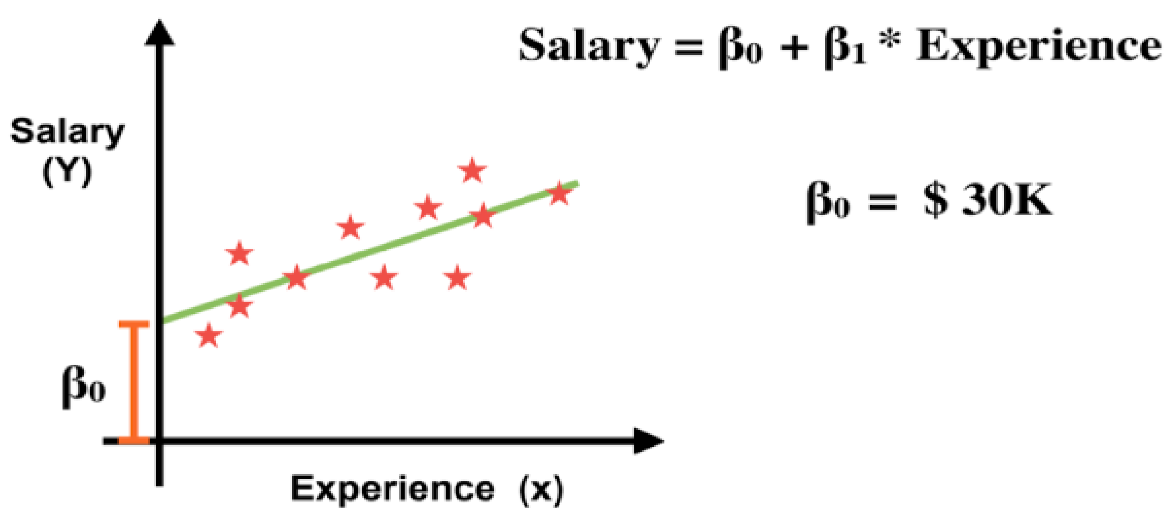
Types of Regression Algorithms:

- Linear regression
- Multiple linear regression
- Polynomial regression
- Ridge regression
- Lasso regression
- ElasticNet regression

Linear Regression

Linear Regression is a statistical model used to predict the relationship between independent and dependent variables denoted by x and y respectively

Simple Linear Regression is where only one independent variable is present and the model has to find the linear relationship of it with the dependent variable



$$Y = mx + c$$

Performance Metrics for Regression Problem

Error = Y (actual) — Y (predicted)

Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

> *Where,*
>
> *N = total number of data points*
>
> *Yi = actual value*
>
> *Ŷi = predicted value*

the lower the MAE, the less error in your model.

Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

> *Where,*
>
> *n = total number of data points*
>
> *Yi = actual value*
>
> *Ŷi = predicted value*

Thus, as with MAE, the lower the MSE, the less error in the model.

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

> *Where,*
>
> *n = total number of data points*
>
> *Yi = actual value*
>
> *Ŷi = predicted value*

lower RMSE → lower error.

R-Squared (R^2)

The R^2 metric gives an indication of how well a model fits your data, but is unable to explain if your model is good or not.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

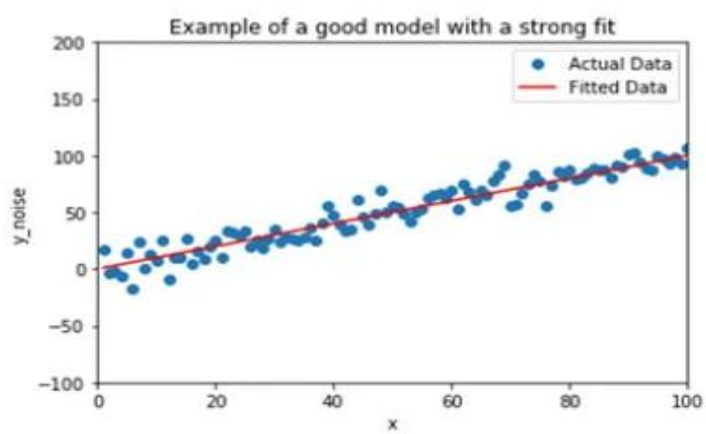
y = dependent variable values, y_hat = predicted values from model, y_bar = the mean of y

The R^2 value ranges from 0 to 1, with higher values denoting a strong fit, and lower values denoting a weak fit. Typically, it's agreed that:

$R^2 < 0.5 \rightarrow$ Weak fit

$0.5 \leq R^2 \leq 0.8 \rightarrow$ Moderate fit

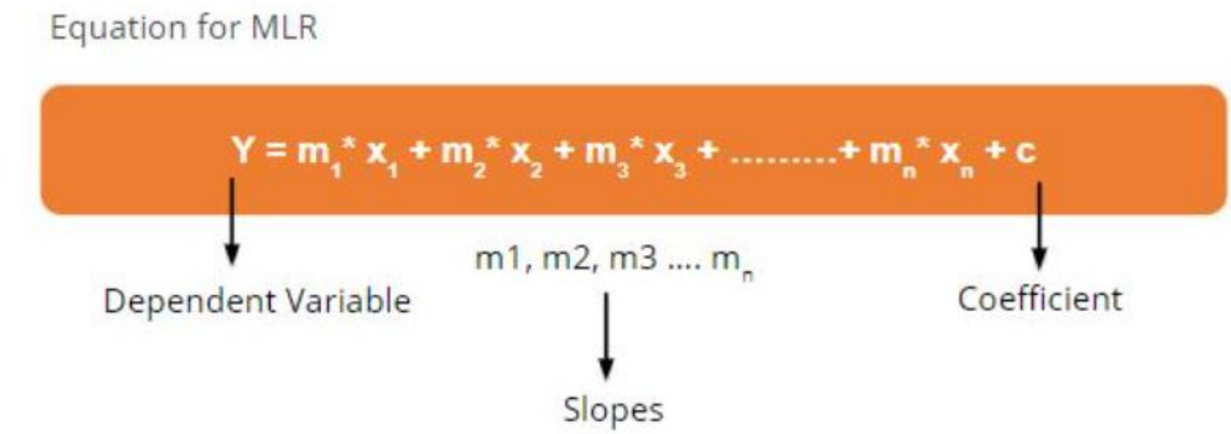
$R^2 > 0.8 \rightarrow$ Strong fit

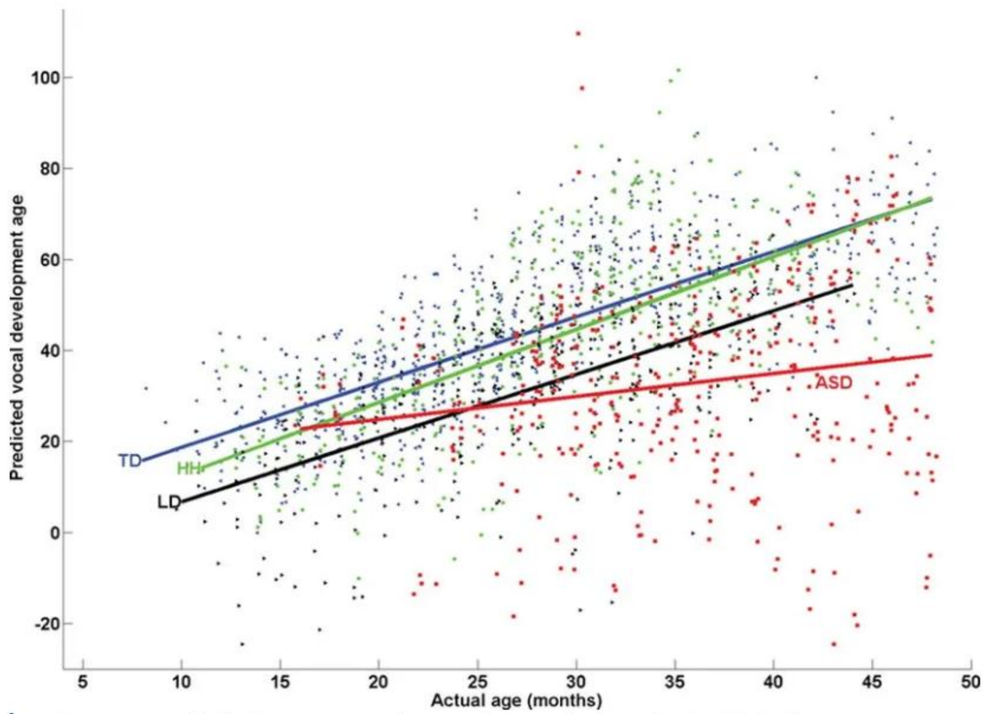



```
R^2 =  0.9186211060937284
```

Muliple Linear Regression:

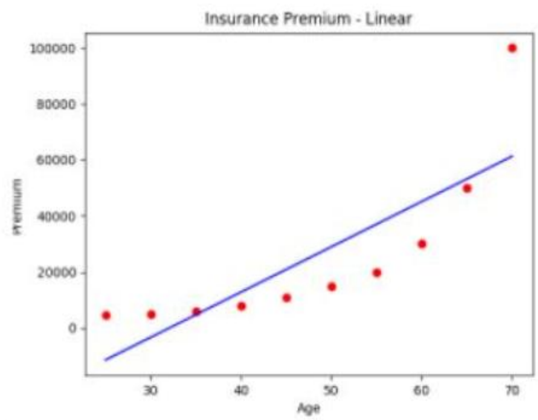
Multiple Linear Regression is one of the important regression algorithms, which models the linear relationship between a single dependent continuous variable and more than one independent variable.

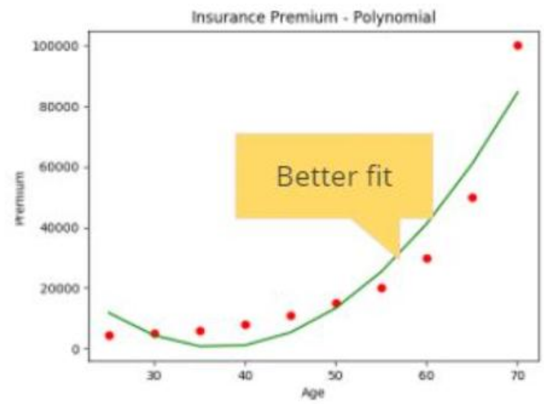


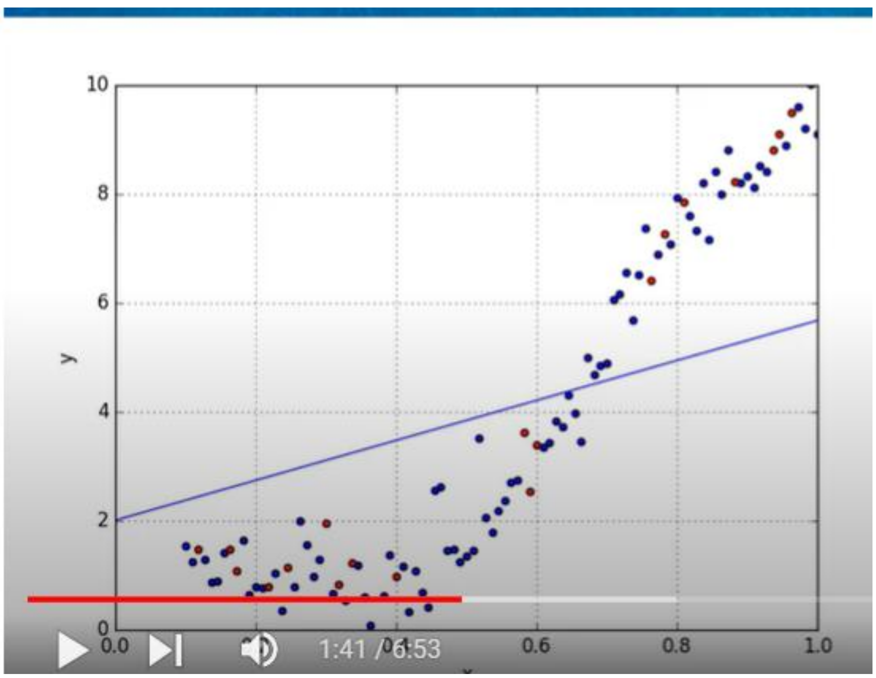


Polynomial Regression:

It is a form of regression analysis in which the relationship between the independent variables and dependent variables are modeled in the **nth degree polynomial**.







Simple Linear Regression

$$y = b_0 + b_1 x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

Polynomial Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ldots + b_n x_1^n$$

Polynomials	Form	Degree	Examples
Linear Polynomial	p(x): ax+b, a $\neq$0	Polynomial with Degree 1	x + 8
Quadratic Polynomial	p(x): ax^2+b+c, a $\neq$ 0	Polynomial with Degree 2	3x^2-4x+7
Cubic Polynomial	p(x): ax^3+bx^2+cx, a $\neq$ 0	Polynomial with Degree 3	2x^3+3x^2+4x+6

It does not require the relationship between the independent and dependent variables to be linear in the data set.