Regression

- In Simple Linear Regression, we assumed that the **data is linearly related to the target** and hence can be modelled with linear models such as linear regression.
- If we find data is more complex & non linear, we use linear model by adding a simple step of **polynomial regression**.
- The high degree polynomial comes with its own cost. The curve may fit the training data very precisely but it may start missing the estimation on test data and this scenario is called as **overfitting.**

- **Overfitting :**
- When the model is able to fit the train data perfectly and the train error is very small.
- But test error is huge which is indicating that the model is unable to predict on unseen data then we say the model has overfitted the data.
- And this happens due to something called as **Low bias (Low Training Error)** and **high variance(High Testing Error)**

Low bias (Low Training Error) & **high variance (High Testing Error)**

- **Underfitting :**

- When the model is unable to fit to the train data and the train error is high (**High bias**) & test error is high **(High Variance)** then we say that the model is underfitting.

High bias (Training Error) & **High variance/Low variance (Testing Error)**

$$Y = 2.3X1+4X2+40X3$$

High val of slope increases complexity of model & computation power also.

Regularization :

- Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting.
- Its a solution to overfitting.
- The solution to underfitting is to train more. Or improve the data.
- Regularization basically modifies the loss function and ensures that overfitting is reduced.
- In Regression problem, regularization is applied using

1) **Ridge Regression**
2) **Lasso Regression**
3) **Elastic Net**

Ridge Regression (L2):

LF/CF = MSE + lambda * (M1^2 + M2^2 + M3^2)

- It penalizes high value of slope.
- Uses when input features are highly correlated with each other.
- Slope value move towards zero.
- The value of lambda varies from (0 to ....)

Lasso Regression (L1):

LF/CF = MSE + lambda * |M1 + M2 + M3|

- It penalizes high value of slope.
- Slope val reaches to zero for few features.

- Helps in feature selection.
- The val of lambda vary from (0 to ….)

Elastic Net Regression:

LF = MSE + lambda * |M1 + M2 + M3| + lambda *(M1^2 + M2^2 + M3^2)

- It combines benefits of ridge & lasso.
- Multicolinarity (High correlation val between independent feature) from Ridge.
- Feature Selection from Lasso.
- The val of lambda vary from (0 to ….)