# Assignment 1

## Report

### Team:

1. Chandan Shrivastava
2. Urvish Pujara

## Task 1 - Linear Regression

sklearn.linear_model is a programme that conducts linear and polynomial regression and provides predictions. Given input and output as inputs, LinearRegression().fit() calculates the best weight values. It returns the model variable itself. It is appropriate for the model. Ordinary least squares Linear Regression is implemented via the model LinearRegression().fit().
A linear regression model is one in which the input variables (x) and the single output variable (y) have a linear relationship (y). That y can be determined using a linear combination of the input variables is more detailed (x). To construct or train the linear regression equation using data, many strategies may be utilised, the most popular of which is termed Ordinary Least Squares. Ordinary Least Squares Linear Regression, or simply Least Squares Regression, is a term used to describe a model created in this manner.
The following is an example of a basic or simple regression problem: y = a + bx
By reducing the sum of the squared residuals, the Ordinary Least Square is used to determine unknown parameters. The total of the squared variances between the observed and fitted values is minimised using this approach, which creates a line between the data points. As a result, the most appropriate coefficient values are provided.

## Task 2 - Calculating Bias and Variance

Bias: The algorithm's tendency to continuously learn the erroneous thing by not taking all of the underfitting data into account is known as bias.

Observation from the table:

1. It is highest at degree 1 and almost the same for degree 2.
2. It is lowest at degree 6.
3. It is almost consistent from degree 3 to degree 10 with some slight changes.
4. It suddenly increases at degrees 11, 12 , then decreases at 13, then increases at 14 and keeps
   fluctuating.

Variance: Variance is an inaccuracy caused by the training set's responsiveness to tiny variations. Because of the high variance, an algorithm may simulate the random noise in the training data instead of the expected overfitting results. The square of the bias should decrease and variance should increase as the degree of the polynomial increases in an ideal condition.

Observation from the table:

1. It is lowest at degree 1.
2. It is highest at degree 15.
3. There is an increase in variance till degree 11.

4. After that, it increases and decreases randomly but overall on increment in degree the variance also gives an increasing trend till degree 15.

# Task 3 - Calculating Irreducible Error

The mistake that cannot be decreased by using appropriate models is known as irreducible error. It's a metric for how much noise there is in our data.

Observation from the table:

1. The irreducible error is approximately zero. At some instance it changes to some degrees around $10^{-10}$ to $10^{-11}$ but they all can be neglected and can be assumed as zero compared to our data points. The zero or negligible irreducible error refers to the fact that no other variable, other than x, has any influence on the output y, or that other variables have only a little effect. The irreducible error does not change when the degree of the polynomial is changed, showing that our model has very low noise.

## Task 4: Plotting Bias$^2$-Variance Graph

Underfitting: It implies that the data does not contain enough data points and that the model does not perform well on training or test data.

1. At low degrees, the model undergoes underfitting which is due to high bias error even though it has low variance.

Overfitting: It signifies that the data closely matches the training data. It doesn't perform well on test data since the model isn't sufficiently broad.

1. Overfitting is indicated by an increased trend in variance.
2. High variance error is responsible for the rising trend in error after a decrease.

Best fit: It signifies the model has a sufficient number of data points for generalisation. It is the polynomial degree that is optimum in this case.

1. The best fit at degree 3 as total error drops sharply at 3 and reaches its minimum.
2. Intersect of Bias$^2$ and variance implies that it is the best fit for our model.

Type of Data: On the basis of the above, observations and data analysis are made as follows:

1. A dataset with a large bias learns quicker and is simpler to comprehend, but it is less versatile. This means that they have a weaker predictive performance, contradicting basic algorithm bias assumptions.
2. The enormous drop in error, which has reached its lowest point, indicates the best fit. The constant increase in error is related to variance error, meaning that at increasing levels of complexity, the model performs unexpectedly or unpredictably due to both variance and bias error.
3. The bias is nearly constant and lowest between degrees 3 and 10, showing that the curve is best suited to the data and that the data is well anticipated, but the rise in variance suggests sensitivity to tiny changes in input values.
4. Following that, the rise in complexity suggests that the model does not correctly contain the training dataset characteristics, and the high variance indicates that the data is vulnerable to exceptions.