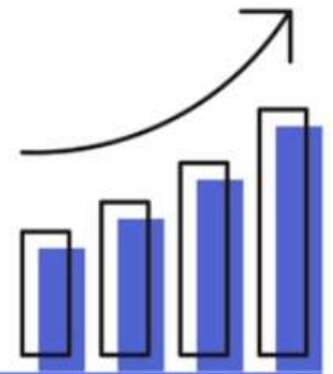




Data Science
Project

Predicting Diamond Sale Price using Neural Network Regression





Learning Outcomes:

In this project we will:

- Learn and understand about Regression
- Learn how to clean and process dataset (outlier, missing values, etc.)
- Learn about Correlation and feature selection.
- Learn about Artificial Neural Networks.
- Deploy our model as a Flask webapp



Part I (i): Regression and Multiple Linear Regression

Regression is a method which measures the nature and extent of the relationship between two or more variables, thus enables us to make predictions.

Regression data exist all around us, and within us as well:

- Teen Birth Rate and Poverty Level Data
- Rainfall measurements
- Stock prices
- Number of sunspots
- Annual retail sales
- Testing automobiles

In this project, we are going to look at such a regression problem, predicting the sale price of a diamond.

✕ Part I (ii): Diamond and its Sale Price

A diamond has various features and characteristics, which are measured and analysed to decide on a market value. Let us look at some of them:

- **Cut:** refers to one of the 10 or so most common diamond cuts.
- **Colour:** Clear diamonds are graded D-Z. The higher letters more yellowish but are often better values since colour is hard to determine once in a ring.
- **Clarity:** refers the inclusions (i.e., internal flaws) in the diamonds seen though a jeweller's loupe or microscope. Fewer and smaller are better.
- **Eye-clean:** refers to the blemishes or inclusions can see with the naked eye. There are 10 grades.

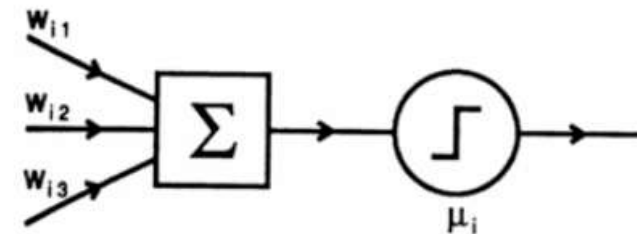


Part I (iii): The Dataset

- Link: <https://www.kaggle.com/datasets/hrokrin/the-largest-diamond-dataset-currely-on-kaggle>
- This dataset has 219703 samples of every kind of diamond
- The task at hand: Predicting the sale price in dollars.
- Let us explore the different features in the dataset and try to understand how we can proceed with model building.

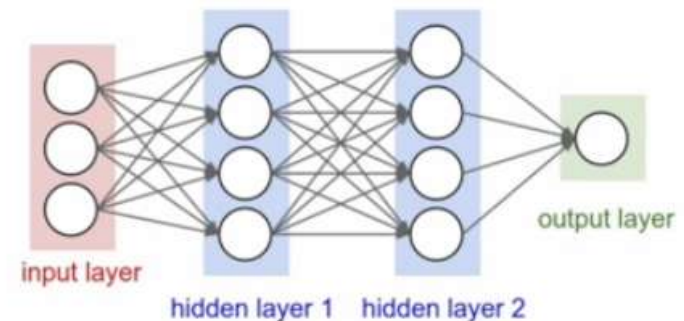
Part II (i): Artificial Neural Networks

- Artificial Neural Networks are essentially an adaptation of the human brain in machines.
- Just like there are neurons firing in the brain to do functions, ANNs also have "neurons" which have some input and output with activation functions. (Fig 1)
- We will be making a Feed-Forward Neural Network.
- Training is done by inputting the data and changing the weights iteration by iteration to get the desired output.
- Small increments are made to the weights to minimize the overall loss.
- Here, by loss, we mean the mean squared error between predicted value and actual value.
- Fig 2 shows the equation of the given neuron.
- Fig 3 shows a Feed-Forward Neural Network with 3 inputs, 2 hidden layers and one output.



$$v = \sum_{j=0}^m w_j x_j$$

$$w_0 = b$$





Part II (ii): Implementing FF-ANN

- We will implementing the Feed-forward Neural network using TensorFlow.
- TensorFlow provides tools to create complex NN architectures and gives the freedom to set and tune hyperparameters.
- Using the feature variables and FF-ANN, we will implement a Regression model
- The model will be then saved and used to predict with test data and then be used to deploy a Flask webapp.