**1. Abstract**

With the increase in the number of graduates who wish to pursue their education, it becomes more challenging to get admission to the students' dream university. Newly graduate students usually are not knowledgeable of the requirements and the procedures of the postgraduate admission and might spent a considerable amount of money to get advice from consultancy organizations to help them identify their admission chances. However, giving the limited number of universities that can be considered by a human consultant, this approach might be bias and inaccurate. Thus, in this paper, a machine learning approach is developed to automatically predict the possibility of postgraduate admission to help graduates recognizing and targeting the universities which are best suitable for their profile.

**2. Introduction**

Machine learning is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence.

In this project, we will be using the Admission\_predict dataset in csv format to predict the chances of students getting admission by a university based on several academic performance measurement. To yield the most accurate result, we will train regression models to find the probability of a student getting accepted into a particular university based on their profile, make prediction and measure its performance.

**3. Problem definition**

To predict the chance of admission of a student based on several academic performance measurement.

**4. Software Used:**

Jupyter Notebook

**5. Dataset information**

This dataset is created for prediction of Graduate Admissions from an Indian perspective.

Content:

The dataset contains several parameters:

1. GRE Scores ( out of 340 )
2. TOEFL Scores ( out of 120 )
3. University Rating ( out of 5 )
4. Statement of Purpose and Letter of Recommendation Strength ( out of 5 )
5. Undergraduate GPA ( out of 10 )
6. Research Experience ( either 0 or 1 )
7. Chance of Admit ( ranging from 0 to 1 ).

The last column is the response variable, which is a continuous value between 0 and 1, indicating the chances of getting admission

**6. About the model**

We have applied the linear regression model to our dataset. Linear regression is perhaps one of the most well known and well understood algorithms in statistics and machine learning.

Machine learning, more specifically the field of predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. In applied machine learning we will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.

As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables, but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm.

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x).

When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression.

Different techniques can be used to prepare or train the linear regression equation from data, the most common of which is called Ordinary Least Squares. It is common to therefore refer to a model prepared this way as Ordinary Least Squares Linear Regression or just Least Squares Regression.

**6.1 Gradient Descent**

When there are one or more inputs you can use a process of optimizing the values of the coefficients by iteratively minimizing the error of the model on your training data.

This operation is called Gradient Descent and works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

When using this method, you must select a learning rate (alpha) parameter that determines the size of the improvement step to take on each iteration of the procedure.

Gradient descent is often taught using a linear regression model because it is relatively straightforward to understand. In practice, it is useful when you have a very large dataset either in the number of rows or the number of columns that may not fit into memory.

*Step 1*

Initialize values β0​, β1​,…, βn​ with some value. In this case we will initialize with 0.

*Step 2*

Pick a value for the learning rate α. The learning rate determines how big the step would be on each iteration.

* If α is very small, it would take long time to converge and become computationally expensive.
* If α is large, it may fail to converge and overshoot the minimum.

Therefore, plot the cost function against different values of α and pick the value of α that is right before the first value that didn’t converge so that we would have a very fast learning algorithm that converges.

Diagram

Description automatically generated

*Step 3*

Make sure to scale the features (X) if it’s on a very different scales. If we don’t scale the data, the level curves (contours) would be narrower and taller which means it would take longer time to converge.

Chart

Description automatically generated

*Step 4*

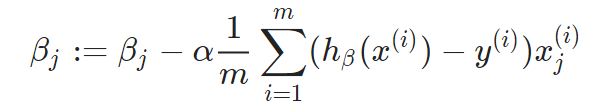
Iteratively update,

A picture containing watch, clock

Description automatically generated

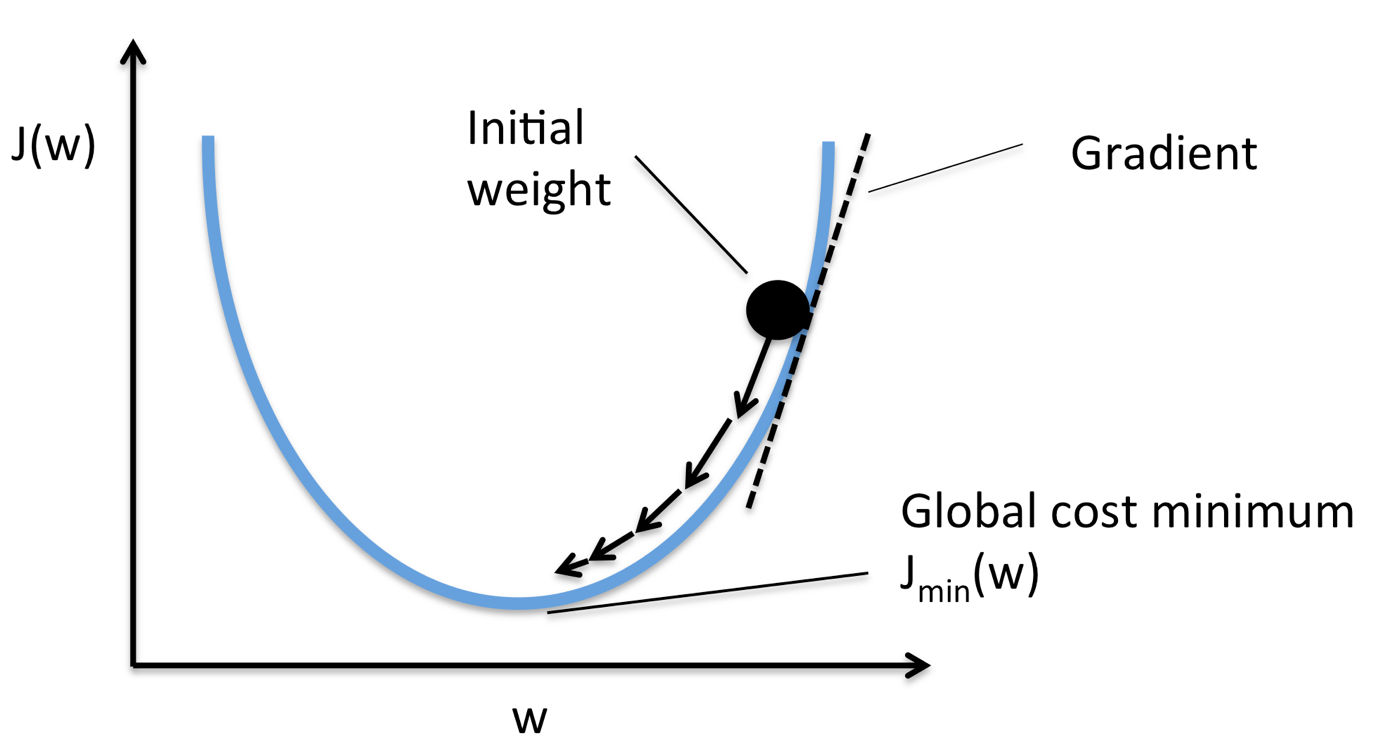
We are changing the values of βj​ in a direction in which it reduces our cost function. And Gradient gives the direction in which we want to move. Finally, we will reach the minima of our cost function. But we don’t want to change values of βj​ drastically, because we might miss the minima. That’s why we need learning rate.

We can’t use above in our mathematical computational model , first we have to apply some magic of math(differentiation of cost).



We get the final formula to update our parameter βj and get the optimal value iteratively.

The algorithm will converge when slope becomes 0.



**6.2 Ordinary Least Square**

When we have more than one input we can use Ordinary Least Squares to estimate the values of the coefficients.

The Ordinary Least Squares procedure seeks to minimize the sum of the squared residuals. This means that given a regression line through the data we calculate the distance from each data point to the regression line, square it, and sum all of the squared errors together. This is the quantity that ordinary least squares seeks to minimize.

This approach treats the data as a matrix and uses linear algebra operations to estimate the optimal values for the coefficients. It means that all of the data must be available and you must have enough memory to fit the data and perform matrix operations.

It is unusual to implement the Ordinary Least Squares procedure yourself unless as an exercise in linear algebra. It is more likely that you will call a procedure in a linear algebra library. This procedure is very fast to calculate.

You will have your features (X) and the target (y). This is how to express the model:

Text

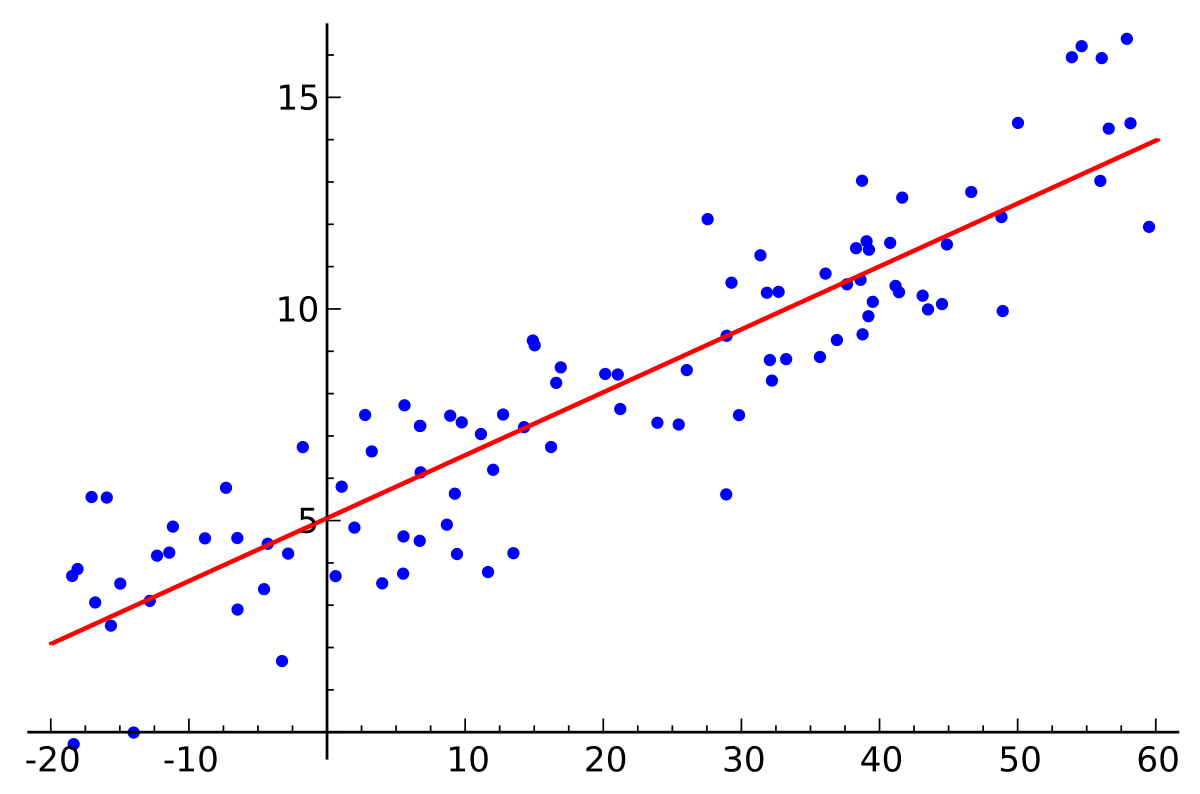
Description automatically generated

Where y is the vector of the target variable, X is a matrix of features, beta is a vector of parameters that you want to estimate, and epsilon is the error term. From the dataset, you’ll want to split features (X) from the target (y), and also add a vector of ones to X for the intercept (or bias) term.

Once done, you can obtain coefficients by the following formula:

Text

Description automatically generated



**6.3 KNN (K-nearest Neighbours Regression)**

KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighbourhood. The size of the neighbourhood needs to be set by the analyst or can be chosen using cross-validation to select the size that minimises the mean-squared error.

A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbours. Another approach uses an inverse distance weighted average of the K nearest neighbours. KNN regression uses the same distance functions as KNN classification.



The above three distance measures are only valid for continuous variables. In the case of categorical variables you must use the Hamming distance, which is a measure of the number of instances in which corresponding symbols are different in two strings of equal length.

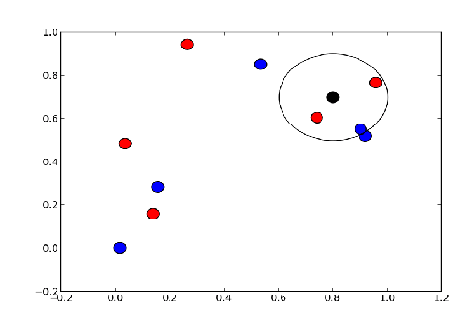
The K-Nearest Neighbor algorithm (KNN) is an elementary but important machine learning algorithm. KNN can be used for both classification and regression predictive problems. The reason for the popularity of KNN can be attributed to its easy interpretation and low calculation time.

KNN uses a similarity metric to determine the nearest neighbors. This similarity metric is more often than not the Euclidean distance between our unknown point and the other points in the dataset.

The general formula for Euclidean distance is:



where q1 to qn represent the attribute values for one observation and p1 to pn represent the attribute values for the other observation.



**7. Results**

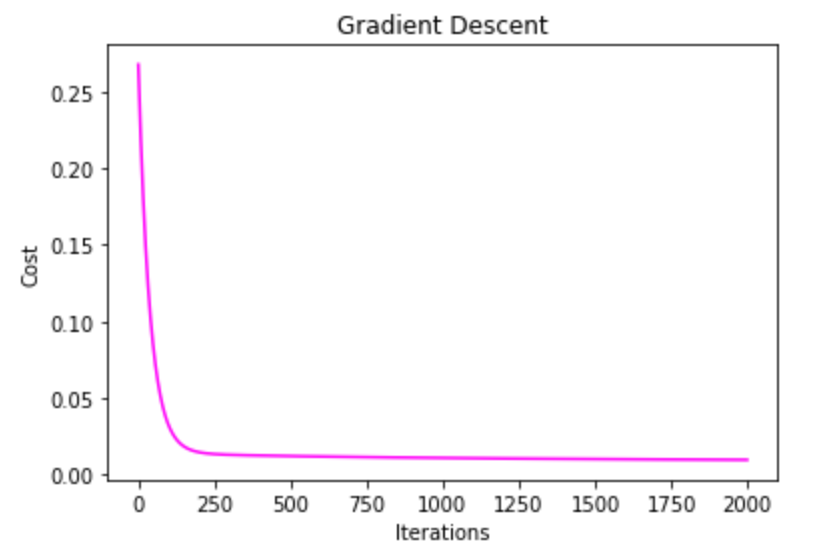
****

Figure 1. Gradient Descent

**Chart, bar chart, histogram

Description automatically generated**

Figure 2. Accuracy for LR model

**Chart, bar chart

Description automatically generated**

Figure 3. Accuracy comparison

**8. Conclusions & Future work**

KNN shows better accuracy than Linear Regression model on our dataset to predict the chance of admission in a university. The accuracy of the LR model is 92.71% whereas accuracy for KNN model is 99.93%.

With the addition of any attributes based on survey data, we can apply this to newly incoming students, but past students are unlikely to provide this data accurately or at all.