Machine Learning: Supervised Learning

Rahul Jain[1], Sriram Gupta Kaluva[2], Sandhya Reddy[3]. Jaya R[4]

[1]*Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore-560103 Karnataka, India*
[2]*Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore-560103 Karnataka, India*
[3]*Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore-560103 Karnataka, India*
[4]*Department of Computer Science and Engineering, New Horizon College of Engineering, Bangalore-560103 Karnataka, India*

[1]vrahul1998@gmail.com
[2]kaluvasriram@gmail.com
[3]sandhyareddyn1998@gmail.com
[4]jayamanojkumar@gmail.com

***Abstract*— This is the age of Artificial Intelligence, machine learning and big data .With the world moving towards all these new technologies it is important to keep oneself updated with all of them. This article will be providing basic information about the workings of supervised learning which is a part of machine learning. It will help get a start in learning more about how machine learning is done and how to implement these techniques in day to day problems.**

***Keywords*—machine learning, linear regression, logistic regression, polynomial regression, random forest**

I. INTRODUCTION

Data is the lifeblood of all business. Data driven decisions increasingly make the difference between keeping up with competition or falling further behind. Machine learning can be the key to unlocking the value of corporate and customer data and enacting decisions that keep a company ahead of the competition.

A subset of artificial intelligence (AI), machine learning (ML) is the area of computational science that focuses on analysing and interpreting patterns and structures in data to enable learning, reasoning, and decision making outside of human interaction. Simply put, machine learning allows the user to feed a computer algorithm an immense amount of data and have the computer analyse and make data-driven recommendations and decisions based on only the input data. If any corrections are identified, the algorithm can incorporate that information to improve its future decision making.

Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

• Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if it's performance on T, as measured by P, improves with experience E.

A. *Working of machine learning*

Machine learning is made up of three parts:

- The computational algorithm at the core of making determinations.
- Variables and features that make up the decision.
- Base knowledge for which the answer is known that enables (trains) the system to learn.

Initially, the model is fed parameter data for which the answer is known. The algorithm is then run, and adjustments are made until the algorithm's output (learning) agrees with the known answer. At this point, increasing amounts of data are input to help the system learn and process higher computational decisions.

According to the prediction of IDC Future escapes, two-thirds of Global 2000 Enterprises CEOs will centre their corporate strategy on digital transformation. A major part of the strategy should include machine-learning (ML) solutions. The implementation of these solutions could change how these enterprises view customer value and internal operating model today.
If you want to stay ahead of the game, then you cannot afford to wait for that to happen. Your digital business needs to move towards automation now while ML technology is developing rapidly. Machine learning algorithms learn from huge amounts of structured and unstructured data, e.g. text, images, video, voice, body language, and facial expressions. By that, it opens a new dimension for machines with limitless applications from healthcare systems to video games and self-driving cars.

B. Machine learning algorithms

Machine learning works on algorithms such as:

1.Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

II. SUPERVISED LEARNING

Supervised learning is said to be a type of learning in ML (Machine Learning) and AI (Artificial Intelligence) where both input and output data is available. The input is used to generate outputs using various algorithms. These outputs are then compared with the actual provided output data to calculate the error. The error is then taken into consideration to re-evaluate the algorithm values to get a more accurate result.

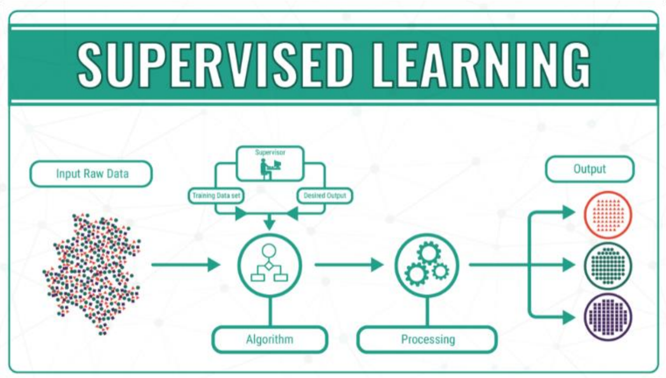

Fig 1 Supervised Learning

A. Examples of Supervised Learning

Some of the examples of supervised learning are:

- Logistic Regression
- Linear Regression

In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships among variables.

IV. LINEAR REGRESSION

In statistics Linear regression is a linear approach of modelling the relation between dependent variable and independent variable (which are also called as scalar response and explanatory variable) by examining two factors.

Linear regression can be statistically/mathematically defined as

$$\hat{y} = \boldsymbol{w}^\top \boldsymbol{x}$$

Fig 2 linear regression equation

where x, y, w are vectors of real numbers and w is a vector of weight parameters. The equation is also written as: y = wx + b, where b is the bias or the value of output for zero input.

A. Interpretation

The fitted linear regression model obtained can be used to identify the relationship between a single predictor variable x and the response variable y when all the other predictor variables in the model are "held fixed".

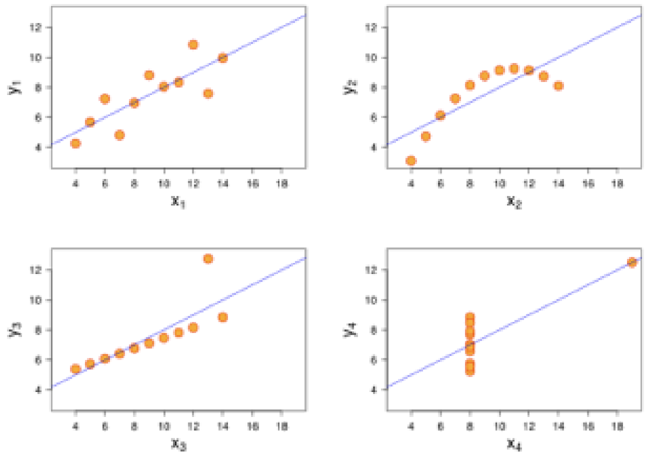

Fig 3. Regression model of Anscombe's quartet

B. Types of linear regression

- simple linear regression: it is the case in which there is only one independent variable.
- multi or multiple linear regression: it is the case in which there are more than one independent variables.

We have one more type of regression which is called as multivariate linear regression in which there are more than one scalar response.

C. Applications of Linear Regression

Few applications of Linear Regression mentioned below are:

- To determine the economic growth of a country or a state in the coming quarter.
- Can also be used to predict the GDP of a country.
- To predict what would be the price of a product in the future.
- To predict the number of runs a player will score in the coming matches.

V.SIMPLE LINEAR REGRESSION

This is the simplest case of linear regression containing a single scalar predictor variable x and a single explanatory variable y.

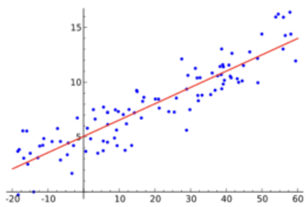

Fig 4. Simple linear regression

The above figure shows the regression model of one independent and one dependent variable.

VI. MULTI LINEAR REGRESSION

This is a regression model which having a vector valued predictor (independent variable) X and one single dependent value is called as multi or multivariable or multiple linear regression.

Mathematically, y = w1x1 + w2x2 + b.

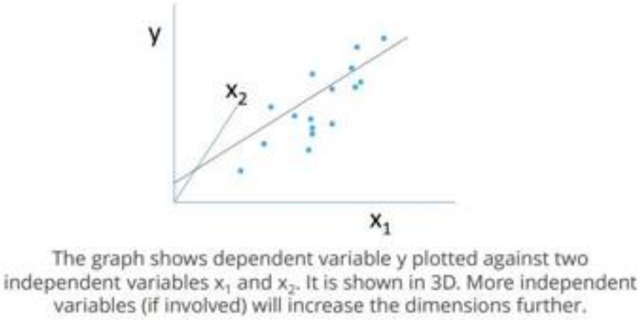

Fig.5 multi linear regression graph

VII. POLYNOMIAL REGRESSION

Polynomial regression is applied when data is not formed in a straight line. It is used to fit a linear model to non-linear data by creating new features from powers of non-linear features.

Example: Quadratic regression

Mathematically,

1. x2 ’ = x2^2
2. y = w1x1 + w2x2^2 + 6 = w1x1 + w2x2 ’ + 6

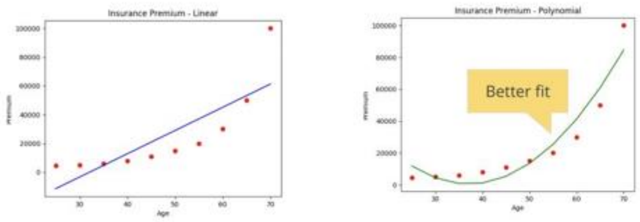

Fig.6 linear and polynomial regression lines

VIII. RANDOM FOREST

It is a process in which the algorithm is used multiple times or a group of different algorithms together to improve the prediction of a model.

It offers no overfitting, resulting less training time. It gives high accuracy even on larger databases. It estimates missing data.

IX. LOGISTIC REGRESSION

Logistic regression is similar to that of linear regression in many ways. It is mainly used for classification problems. Classification problems include classifying whether an email is a spam or not, classifying whether phone number is genuine or not, checking if a website is fraudulent or not.

A. Sigmoid Function

Logistic regression algorithms use a linear equation with independent predictors to predict values. The predicted values can be between positive to negative infinity. The output of the algorithm to a value needs to be 0-no or 1-yes. Therefore, the output of the linear equation should be converted to a value in the range of 0 to 1. To convert the predicted value between 0 and 1, a sigmoid function is used.

Logistic regression is called so for the logistic function used at the core of the method.

The sigmoid function used in logistic regression was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. The curved formed is an S-shaped curve that takes any real-valued number and map it to values between 0 and 1, but never exactly at those limits.

Where e is the base of the natural logarithms (Euler’s number) and the actual numerical value that is to be transformed. Below is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function.

$$z = \theta_0 + \theta_1 \cdot x_1 + \theta \cdot x_2 + \cdots \qquad g(x) = \frac{1}{1+e^{-x}}$$

Linear Equation and Sigmoid Function

$$h = g(z) = \frac{1}{1+e^{-z}}$$

Squashed output-h

Fig 7. Linear equation, sigmoid function and squashed output-h

The output(z) is taken for the linear equation and given to the function g(x) which returns a value h which lies in the range of 0 to 1. To understand how sigmoid function converts the values to values within the range, the graph of the sigmoid function below is shown.

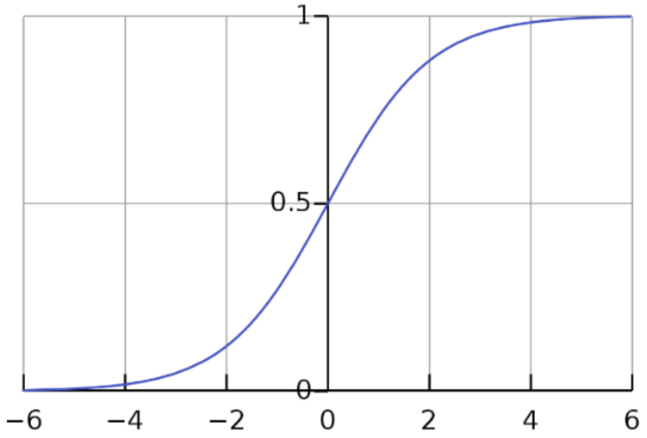

Fig 8. Sigmoid function graph

B. Cost Function

The same cost function in linear regression algorithm cannot be used are we are predicting class values. Therefore, we use a logarithmic loss function to calculate the cost for misclassifying.

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Fig 9. Cost function

The above cost function can be rewritten as below. [10]

$$-\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

Fig 10. Cost Function

C. Calculating Gradients

The partial derivatives of the cost function is taken with respect to each parameter(theta_0,...) to obtain the gradients. Using these gradients the values of theta_,... can be updated.

$$J = \frac{-1}{m} \cdot [\sum_{i=1}^{m} y_i \cdot \log h_i + (1 - y_i) \cdot \log 1 - h_i]$$

$$\frac{\partial J}{\partial \theta_n} = \frac{-1}{m} \cdot [\sum_{i=1}^{m} \frac{y_i}{h_i} \cdot h_i^2 \cdot x_n \cdot \frac{1 - h_i}{h_i} + \frac{1 - y_i}{1 - h_i} \cdot -h_i^2 \cdot x_n \cdot \frac{1 - h_i}{h_i}]$$

$$\frac{\partial J}{\partial \theta_n} = \frac{-1}{m} \cdot [\sum_{i=1}^{m} x_n \cdot (1 - h_i) \cdot y_i - x_n \cdot h_i \cdot (1 - y_i)]$$

$$\frac{\partial J}{\partial \theta_n} = \frac{1}{m} \cdot x_i \cdot [\sum_{i=1}^{m} h_i - y_i]$$

Fig 11. Gradients

IV.CONCLUSION

This journal contains basic details of machine learning and various algorithms in machine learning. It gives further information about supervised learning and the various examples of supervised learning. This journal does not contain complex information that may and may not be understood. It contains all the information required to get a flavour or understanding of how regression is used in machine learning.

REFERENCES

[1] Ethem Alpaydin, Introduction To Machine Learning, Third Edition.

[2] Rohith Gandhi, Introduction to Machine Learning Algorithms: Linear Regression, URL:-https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

[3] Rohith Gandhi, Introduction to Machine Learning Algorithms: Logistic Regression URL:-https://hackernoon.com/introduction-to-machine-learning-algorithms-logistic-regression-cbdd82d81a36