**Machine Learning** is the field of study that gives computers the capability to learn without being explicitly programmed.

**Types of machine learning problems**

**Supervised learning:** The computer is presented with example inputs and their desired outputs, given by a “teacher”, and the goal is to learn a general rule that maps inputs to outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Some real life examples are:

Image Classification: You train with images/labels. Then in the future you give a new image expecting that the computer will recognize the new object.

Market Prediction/Regression: You train the computer with historical market data and ask the computer to predict the new price in the future.

Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

Y = f(X)

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

Supervised learning problems can be further grouped into regression and classification problems.

• Classification: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.

• Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Some common types of problems built on top of classification and regression include recommendation and time series prediction respectively.

**supervised machine learning algorithms are:**

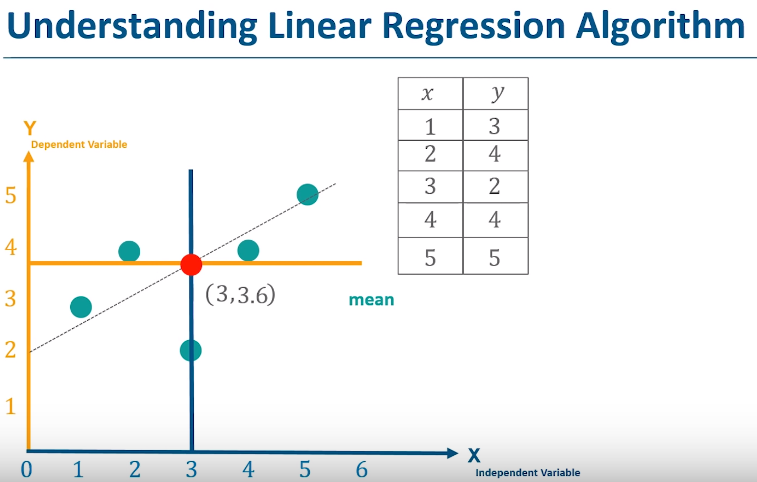
1. Linear regression:-

Mathematical approach:

Plot the graph for the respective x and y values.

Calculate the mean of x and y.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | X- | Y- | (X-)² | (X-)( Y-) |
| ­­1 | 3 | -2 | -0.6 | 4 | 1.2 |
| 2 | 4 | -1 | 0.4 | 1 | -0.4 |
| 3 | 2 | 0 | -1.6 | 0 | 0 |
| 4 | 4 | 1 | 0.4 | 1 | 0.4 |
| 5 | 5 | 2 | 1.4 | 4 | 2.8 |



Here, and are the mean.

Y = mx + c

M = £(x-)( Y-/£ (X-)²

Thus, m = 0.4

Now, and y are same ie, 3.6. X and are same ie, 3

3.6 = 0.4 x 3 + c

C = 3.6 – 1.2 = 2.4

Finally, m = 0.4, c = 2.4. Let’s predict values of y for x = { 1,2,3,4,5 }

Here, y1, y2,y3,y4 and y5 are the prediction values.

Y1 = 0.4 \* 1 + 2.4 = 2.8

Y2 = 0.4 \* 2 + 2.4 = 3.2

Y3 = 0.4 \* 3 + 2.4 = 3.6

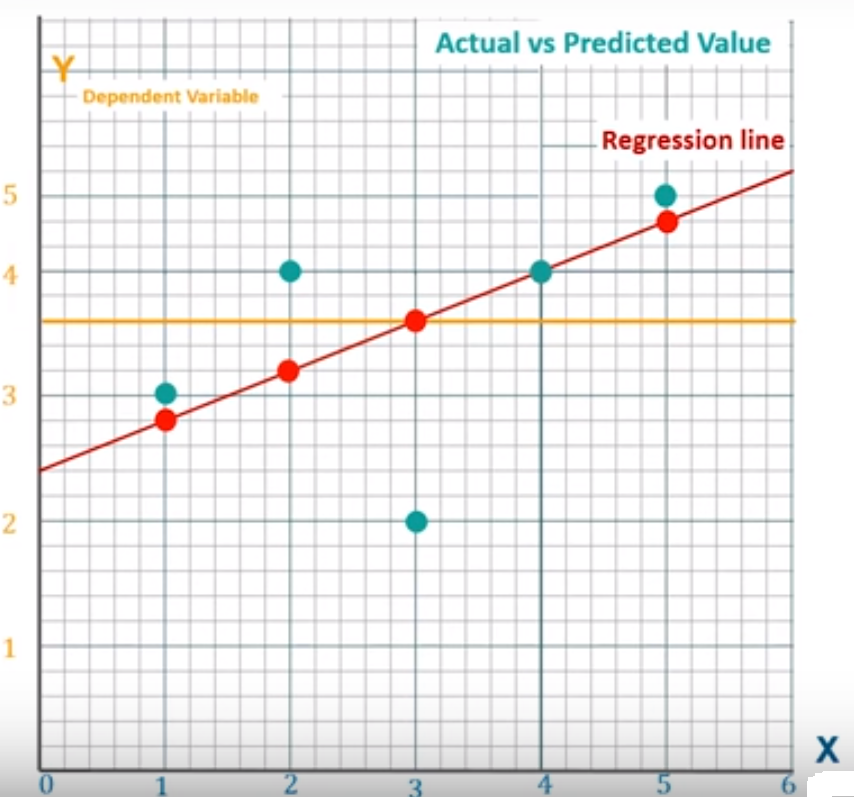
Y4 = 0.4 \* 4 + 2.4 = 4.0

Y5 = 0.4 \* 5 + 2.4 = 4.4

Now, plot the graph for (xi, yi) for above values and it will become the regression line.

R2 Method:- This is to calculate our regression line with actual data. It should be between 0 to 1. More fit regression line should be near to 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Y | Yp | Y- | Yp- | ( Y-)2 | (Yp - )2 |
| ­­1 | 3 | 2.8 | -0.6 | 2.8-3.6=-0.8 | 0.36 | 0.64 |
| 2 | 4 | 3.2 | 0.4 | -0.4 | 0.16 | 0.16 |
| 3 | 2 | 3.6 | -1.6 | 0 | 2.56 | 0 |
| 4 | 4 | 4 | 0.4 | 0.4 | 0.16 | 0.16 |
| 5 | 5 | 4.4 | 1.4 | 0.8 | 1.96 | 0.64 |



R2 = £(Yp - )2/£( Y-)2

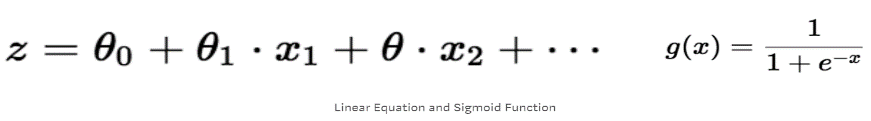
It is approx.. 0.64.

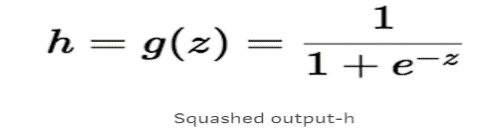
Here, the target value is always numeric.eg, if we have data of the compamy’s profit for every month, then we can predict the company’s profit for over a certain span of time period.

1. Logistic Regression:-

It is the Correlation as the statistical technique that can show whether and how strongly pairs of variables are related.The main result of a correlation is called the correlation coefficient (or "r"). It ranges from -1.0 to +1.0. The closer r is to +1 or -1, the more closely the two variables are related. If r is close to 0, it means there is no relationship between the variables. If r is positive, it means that as one variable gets larger the other gets larger. If r is negative it means that as one gets larger, the other gets smaller (often called an "inverse" correlation). The target value for this algorithm is a categorical value i.e, yes or no.

Logistic regression algorithm also uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. We need the output of the algorithm to be class variable, i.e 0-no, 1-yes. Therefore, we are squashing the output of the linear equation into a range of [0,1]. To squash the predicted value between 0 and 1, we use the sigmoid function.



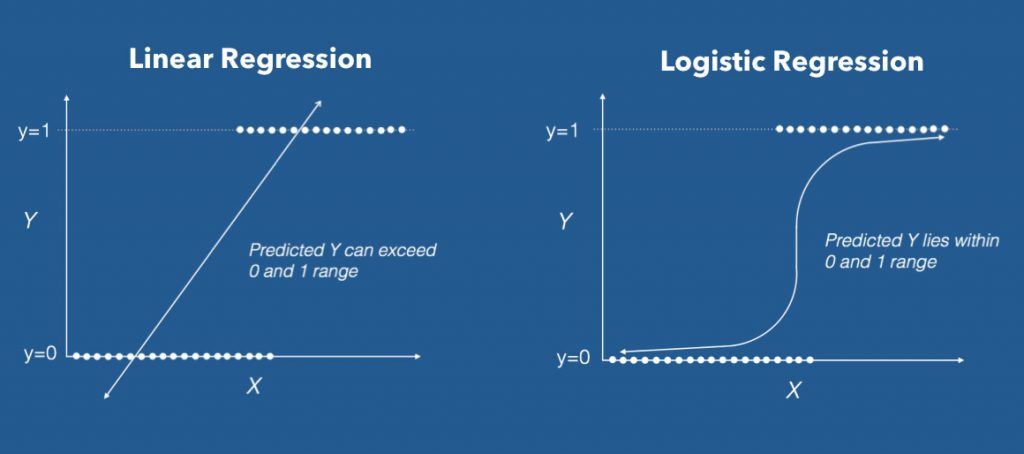


Real world examples of logistic regression includes,

* **Spam Detection** : Predicting if an email is Spam or not
* **Credit Card Fraud** : Predicting if a given credit card transaction is fraud or not
* **Health** : Predicting if a given mass of tissue is benign or malignant
* **Marketing** : Predicting if a given user will buy an insurance product or not
* **Banking** : Predicting if a customer will default on a loan.

Why not linear regression?

When the response variable has only 2 possible values, it is desirable to have a model that predicts the value either as 0 or 1 or as a probability score that ranges between 0 and 1. Linear regression does *not* have this capability. Because, If you use linear regression to model a binary response variable, the resulting model may not restrict the predicted Y values within 0 and 1.

[](https://www.machinelearningplus.com/wp-content/uploads/2017/09/linear_vs_logistic_regression.jpg)

LINEAR VS LOGISTIC REGRESSION

Logistic is categorical means yes or no. linear regression is continuous means some number. This is where logistic regression comes into play. In logistic regression, you get a probability score that reflects the probability of the occurence of the event. An event in this case is each row of the training dataset. It could be something like classifying if a given email is spam, or mass of cell is malignant or a user will buy a product and so on.

The fundamental equation of generalized linear model(glm) is:

g(E(y)) = α + βx1 + γx2

Here, g() is the link function, E(y) is the expectation of target variable and α + βx1 + γx2 is the linear predictor ( α,β,γ to be predicted). The role of link function is to ‘link’ the expectation of y to linear predictor.

Important Points

1. GLM does not assume a linear relationship between dependent and independent variables. However, it assumes a linear relationship between link function and independent variables in logit model.
2. The dependent variable need not to be normally distributed.
3. It does not uses OLS (Ordinary Least Square) for parameter estimation. Instead, it uses maximum likelihood estimation (MLE).
4. Errors need to be independent but not normally distributed.

Let’s understand it further using an example:

We are provided a sample of 1000 customers. We need to predict the probability whether a customer will buy (y) a particular magazine or not. As you can see, we’ve a categorical outcome variable, we’ll use logistic regression.

To start with logistic regression, I’ll first write the simple linear regression equation with dependent variable enclosed in a link function:

       g(y) = βo + β(Age)         ---- (a)

Note: For ease of understanding, I’ve considered ‘Age’ as independent variable.

In logistic regression, we are only concerned about the probability of outcome dependent variable ( success or failure). As described above, g() is the link function. This function is established using two things: Probability of Success(p) and Probability of Failure(1-p). p should meet following criteria:

1. It must always be positive (since p >= 0)
2. It must always be less than equals to 1 (since p <= 1)

Now, we’ll simply satisfy these 2 conditions and get to the core of logistic regression. To establish link function, we’ll denote g() with ‘p’ initially and eventually end up deriving this function.

Since probability must always be positive, we’ll put the linear equation in exponential form. For any value of slope and dependent variable, exponent of this equation will never be negative.

p = exp(βo + β(Age)) = e^(βo + β(Age))    ------- (b)

To make the probability less than 1, we must divide p by a number greater than p. This can simply be done by:

p  =  exp(βo + β(Age)) / exp(βo + β(Age)) + 1   =   e^(βo + β(Age)) / e^(βo + β(Age)) + 1    ----- (c)

Using (a), (b) and (c), we can redefine the probability as:

              p = e^y/ 1 + e^y           --- (d)

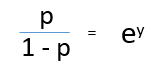
*where*p is the probability of success.*This (d) is the Logit Function*

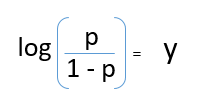
If p is the probability of success, 1-p will be the probability of failure which can be written as:

q = 1 - p = 1 - (e^y/ 1 + e^y)    --- (e)

*where* q is the probability of failure

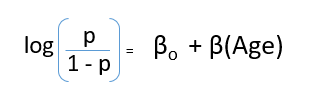
On dividing, (d) / (e), we get,

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/11/1.png)

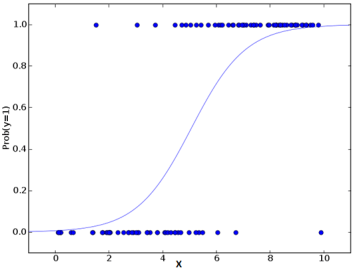
After taking log on both side, we get,  
[](https://www.analyticsvidhya.com/wp-content/uploads/2015/11/2.png)

log(p/1-p) is the link function. Logarithmic transformation on the outcome variable allows us to model a non-linear association in a linear way.

After substituting value of y, we’ll get:

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/11/3.png)

This is the equation used in Logistic Regression. Here (p/1-p) is the odd ratio. Whenever the log of odd ratio is found to be positive, the probability of success is always more than 50%. A typical logistic model plot is shown below. You can see probability never goes below 0 and above 1.

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/11/plot.png)

1. KNN:-

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry.

KNN is a typical example of a lazy learner. It is called lazy not because of its apparent simplicity, but because it doesn't learn a discriminative function from the training data but memorizes the training dataset instead.

We can implement a KNN model by following the below steps:

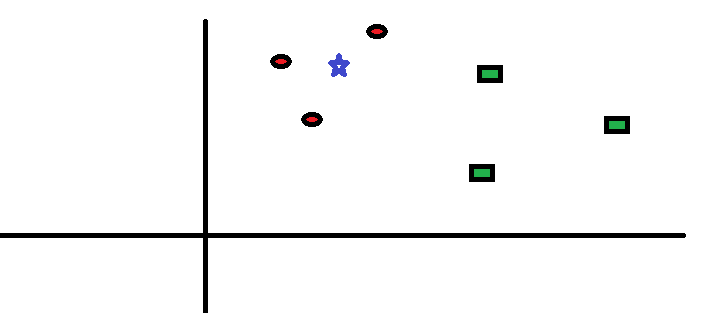
1. Load the data
2. Initialise the value of k
3. For getting the predicted class, iterate from 1 to total number of training data points

* Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it’s the most popular method.
* The other metrics that can be used are Chebyshev, cosine, etc.Sort the calculated distances in ascending order based on distance values.Get top k rows from the sorted array.
* Get the most frequent class of these rows.Return the predicted class

K are the num of neighbours. Biggest use is the recommendation system. Also helps in the searching similar topics in internet as there are thousands of data in internet. Hand writing detection is the advance use.

KNN are supervised learning and also a lazy learner because it does not learn a discriminative func from the training data but memorizes the training dataset instead. Prediction in KNN is relatively expensive. Each time we want to make a prediction, KNN is searching for the nearest neighbours in entire training data set. We use Euclidian distance formula i.e, √(x1-x0)2(y1-y0)2.

Let’s take a simple case to understand this algorithm. Following is a spread of red circles (RC) and green squares (GS) :

[](https://www.analyticsvidhya.com/blog/wp-content/uploads/2014/10/scenario1.png)

You intend to find out the class of the blue star (BS) . BS can either be RC or GS and nothing else. The “K” is KNN algorithm is the nearest neighbors we wish to take vote from. Let’s say K = 3. Hence, we will now make a circle with BS as center just as big as to enclose only three datapoints on the plane. Refer to following diagram for more details:

[](https://www.analyticsvidhya.com/blog/wp-content/uploads/2014/10/scenario2.png)The

three closest points to BS is all RC. Hence, with good confidence level we can say that the BS should belong to the class RC. Here, the choice became very obvious as all three votes from the closest neighbor went to RC. The choice of the parameter K is very crucial in this algorithm