1.

A Linear Separability is a type of neural network that classifies the pattern and represent x and o symbol recognition.

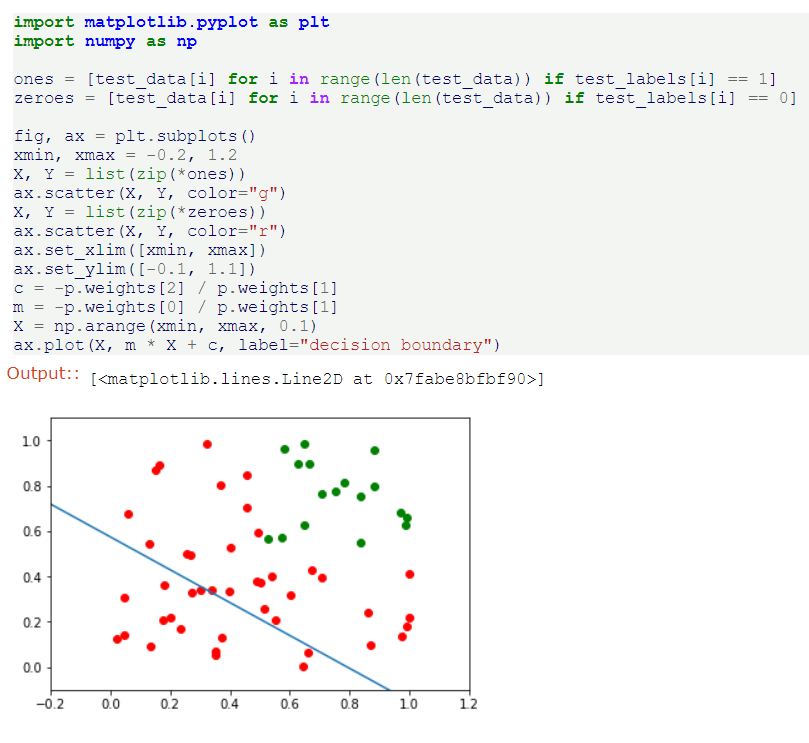
The process unit of a single layer perceptron is able to recognize the pattern where in each network is come to be the classified intent and their working has to be in classification so that they can provide the meaningful information out of the raw data. Each Linear classes has to do the two classes so that it can distinguish between the meaningful data.

Linear separability separates the data between two distinct classes and within the classes it performs XOR gate operations. It refers to the class of pattern with n-dimensional vector x=(x1,x2,....xn)

It differentiates between the red dots and blue dots wherein it can manage the data in linear fashion way.

Below are the attached image of graph of linear and not linear graph.

Linear Graph :



2.

Linear regression produces its output using the sigmoid function and which takes the probability of two or more discrete classes so that it can distinguish about the linear and their behavior of separating two or more classifications.

The Sigmoid means the S shape and also defined as squashing function and the main function is to give the conditional probability and its derivative in the simple form.

The main difference is that it provides the conditional probability and according to that it provides the solution of any discrete classes as it belongs to the logistic regression.

3.

SVM stands for Support Vector Machine and main perspective of SVM machine is to find the hyperplane in present dimensional space in a n number. The plane that has the maximum margin is to be chosen as a hyper plane and the plane is further called in a graph is Optimal hyperplane.

In SVM there are two margins small and large margins. Small margins are described where the distance between both the classification is small and where the classification distance is large then it is called as Large Margins.

The Classification from both the discrete classes touches its separation lines and those classification are known as the support Vectors and those points are known as data points.

It then recognizes the correct plane for the classification of the data points

The term meaningful it is just because that it is giving a very relevant data which can sort many of the important classification which from human perspective it is going to take a lot of time , with the machine learning SVM comes out to be very important model that it can provide distinguish between a data points of large and small length margin and the data points.

Although those were the positive points , but sometimes the SVM also do the meaningless by comparing the 77% of data to the useless datasets wherein it tries to find out the positive data points and data collection..

4.

It first appears to be genuine mathematical sorcery, not to mention the problem of lexical ambiguity (does kernel refer to: a non-parametric way to estimate a probability density, the set of vectors v for which a linear transformation T maps to the zero vector — i.e. T(v) = 0 (linear algebra), the set of elements in a group G that are mapped to the identity element by a homomorphism between groups (group theory), the core of a computer operating system (computer science), or something to do with the seeds of nuts or fruit?).

For practical reasons, it is important to understand because implementing support vector classifiers requires specifying a kernel function, and there are not established, general rules to know what kernel will work best for your particular data.

The kernel trick also illustrates some fundamental ideas about different ways to represent data and how machine learning algorithms “see” these different data representations. And finally, the seeming mathematical sleight of hand in the kernel trick just begs one to further explore what it actually means.

5.

The standard SVM classifier works only if you have a well separated categories. To be more specific, they need to be linearly separable. It means there exist a such that all points belonging to a single category are either below or above it. In many cases that condition is not satisfied, but still the two classes are pretty much separated except some small training data where the two categories overlap. It wouldn’t be a huge error if we would draw a line and accept some level of error - having training data on the wrong side of the marginal hyperplanes. How do we measure the error? The answer is: slack variables.

For each training data point we can define a variable that measures the distance of the point to its marginal hyperplane, lets call it ξ∗iξi∗. Whenever the point is on the wrong site of the marginal hyperplane we quantify the amount of error by the ratio between ξ∗iξi∗ and half of the margin, i.e. distance between separating hyperplane and marginal hyperplane. Points on the correct site are not quantified as errors. This is a geometrical interpretation of slack variables ξiξi. You can now go back to the initial SVM problem and maximize the margin in the presence of errors. The larger the error that you allow for, the wider the margin (numerical illustration at the end).