Support Vector Machine(SVM) is a supervised machine learning model which can be used for both classification (Support Vector Classification[SVC]) and regression (Support Vector Regression[SVR]).

**SVM as a Classifier (SVC):**

It is a discriminative and non-probabilistic classifier which can be used for classifying both the Linearly Separable Dataset and also the Non-Linearly Separable Dataset.

It partitions a *feature space* into different groups, which in our case means separating a collection of articles into different categories.

SVM achieves this by finding an optimal means of separating such groups based on their known categories. It takes the data points and outputs the decision boundary that best separates the categories.

This best decision boundary is the one that maximizes the margins between any two categories. In other words, the decision boundary whose distance to the nearest element of each category is the largest. This best decision boundary is called the **Maximum Margin Classifier** and the nearest points which help in finding this best decision boundary are called the **Support Vectors.** This is the reason that this model(SVM) is called Support Vector Machine.

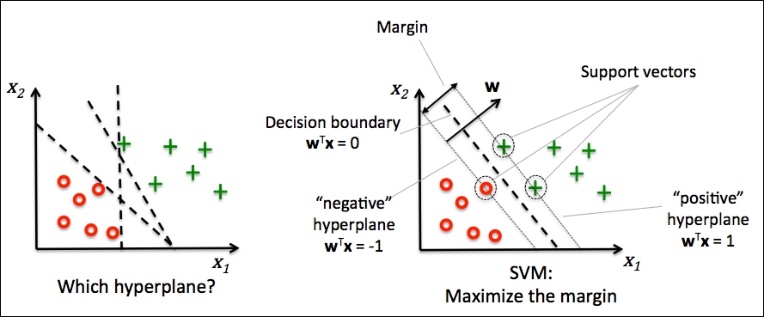
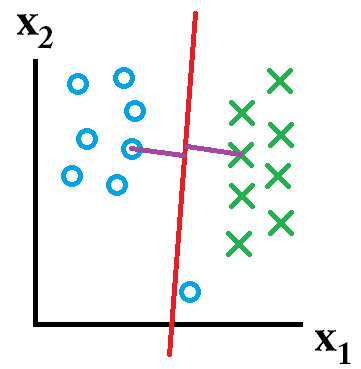
 

Fig.1 Maximum Margin Classifier and Support Vectors in SVM Fig.2 Outlier ignored in general SVM

Only these **Support Vectors** contribute in finding the decision boundaries for the different categories, all other points/vectors don’t contribute anything to the model.

SVM looks at the very extreme case for finding the best boundaries, that is why SVM is very special and very different than most of the other Machine Learning models.

Note that the general SVM does not include **outliers** inside the decision boundary of its category.

But we can change the parameters of SVM to change the way a decision boundary is selected and can find the parameters which are best suited for our application.

Linearly Separable Dataset:

Linearly Separable dataset can be easily classified using the above approach.

Non-Linearly Separable Dataset:

For Non-Linearly Separable dataset, we cannot just simply separate the different categories using a simple line or hyperplane.

Fig 3. Non-linearly separable data Fig 4. Mapping into higher dimension (adding z-axis component)

Z = x^2 + y^2

We are required to take our non-linearly separable dataset, map it to a higher dimension to make it linearly separable, build decision boundaries using SVM and then project all of that back into our original dimensions.

But calculating the transformations to a higher dimensional space can be highly compute intensive and might require a lot of computation and processing power.

So, to overcome this, we use a trick known as **the kernel trick**:

SVM does not need the actual vectors to work on it, it can do it only with the **dot products** between them. This dot product is called a **kernel function**.

This function will save us a lot of expensive calculations.

There are many kernel functions(kernels), some commonly used kernels are:

* Gaussian RBF kernel
* Linear kernel
* Sigmoid kernel
* Polynomial kernel

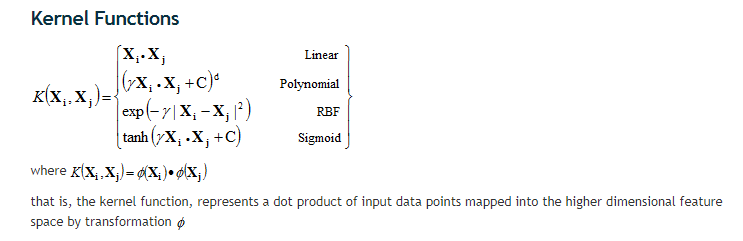


Fig 5. Different types of Kernel functions

Sources:

Fig 1: <https://www.packtpub.com/graphics/9781783555130/graphics/3547_03_07.jpg>

Fig 2: <https://stackoverflow.com/>

Fig 3: <https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_8.png>

Fig 4: <https://www.analyticsvidhya.com/wp-content/uploads/2015/10/SVM_9.png>

Fig 5: <http://www.statsoft.com/textbook/support-vector-machines>