

Author: Kazi Amit Hasan

Department of Computer Science & Engineering,

Rajshahi University of Engineering & Technology (RUET)

Website: <https://amithasanshuvo.github.io/>

ResearchGate: https://www.researchgate.net/profile/Kazi_Amit_Hasan

Linkedin: <https://www.linkedin.com/in/kazi-amit-hasan/>

Email: kaziamithasan89@gmail.com

Support Vector Machine Machine Learning Algorithm

Support vector machine (SVM) is a well known supervised ML algorithm that used for classification and regression both cases. This is based of finding the hyperplane that best divides two classes.

If we break down the term Support Vector Machine, then we will get three terms.
Let's talk about support and vector!

Support vectors are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set.

Another term that came here is 'hyperplane'. Let's know about it!

Hyperplane: As a simple example, for a classification task with only two features (like the image above), you can think of a hyperplane as a line that linearly separates and classifies a set of data.

Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it. So when new testing data is added, whatever side of the hyperplane it lands will decide the class that we assign to it.

How do we find the right hyperplane?

The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly.

There are some advantages and disadvantages of SVM. They are:

Advantages:

1. Accuracy
2. Works well on smaller cleaner datasets
3. It can be more efficient because it uses a subset of training points

Disadvantages:

1. Isn't suited to larger datasets as the training time with SVMs can be high
2. Less effective on noisier datasets with overlapping classes