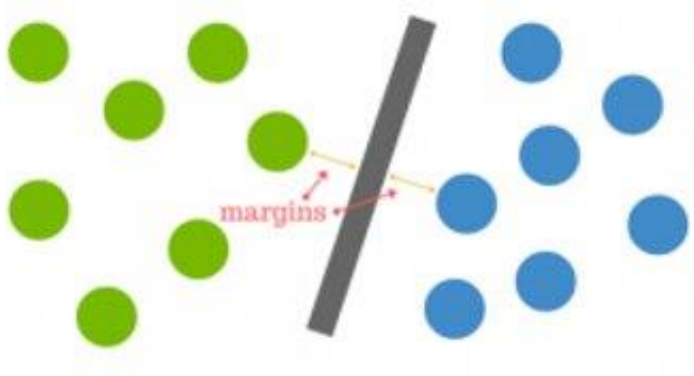
Support Vector Machine

NASTARAN SHAHPARIAN |SHARCNET | COMPUTE ONTARIO | COMPUTE CANADA |YORK UNIVERSITY

Definition

- The original SVM algorithm was developed by Vladimir N Vapnik and Alexey Ya. Chervonenkis in 1963
- Support Vector Machines are machine learning algorithms that are used for classification and regression, and outlier detection.
- SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes, as shown in the image below.



https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

Concepts

Support Vectors

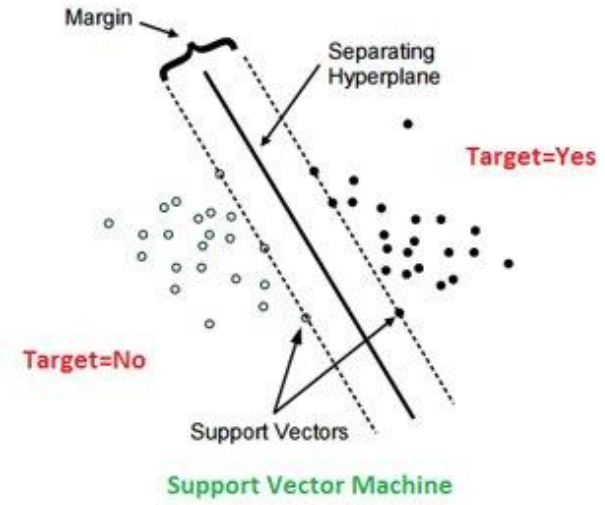
- Data points of the dataset, nearest to the hyperplane

Hyperplane

- A line that linearly separates and classifies a set of data
- The further from the hyperplane data points lie, the more confident we are that they have been correctly classified

Margin

- The distance between the hyperplane and the nearest data point from either set

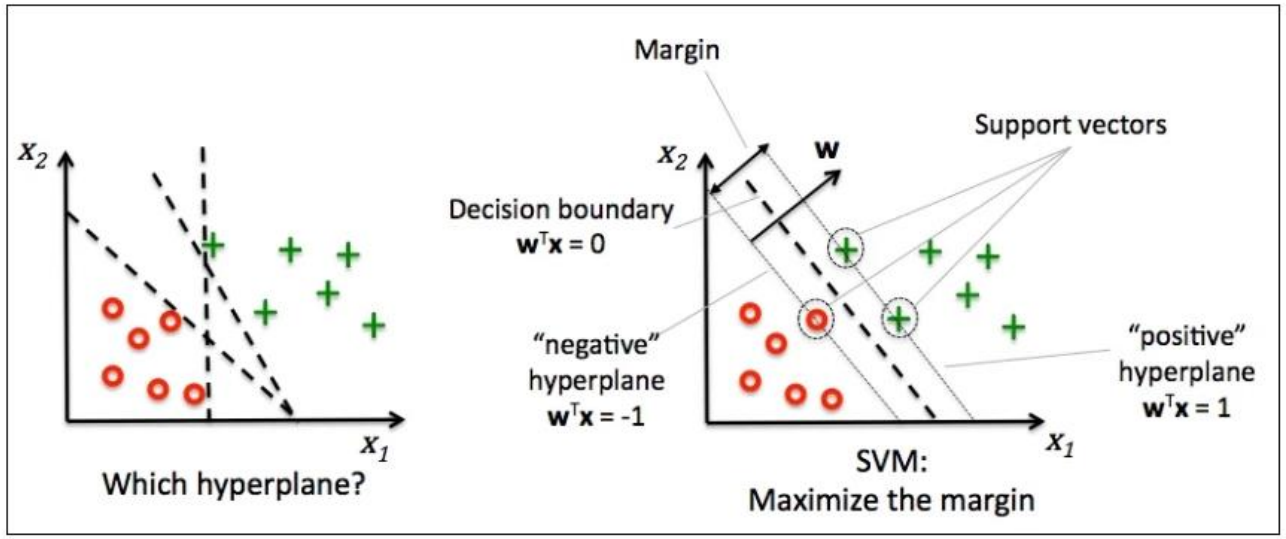


https://www.kaggle.com/code/prashant111/svm-classifier-tutorial

SVM Under the hood

Main objective is to select a hyperplane with the maximum possible margin

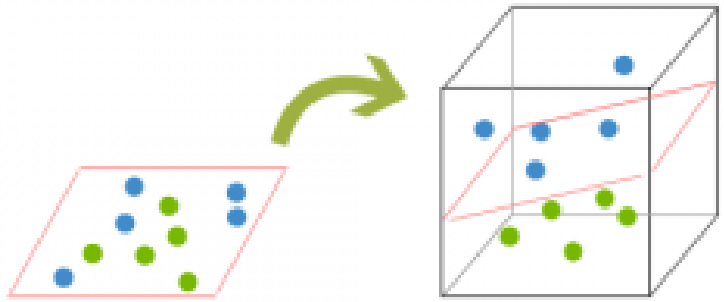
1. Generate hyperplanes which segregates the classes in the best possible way. There are many hyperplanes that might classify the data
2. So, we choose the hyperplane so that distance from it to the support vectors on each side is maximized.



https://www.kaggle.com/code/prashant111/svm-classifier-tutorial

What if data isn't linearly separable?

- It's necessary to move away from a 2d view of the data to a 3d view
- Mapping of data into a higher dimension. This is known as kernelling
- Hyperplane can no longer be a line. It must now be a plane
- The data will continue to be mapped into higher and higher dimensions until a hyperplane can be formed to segregate it.

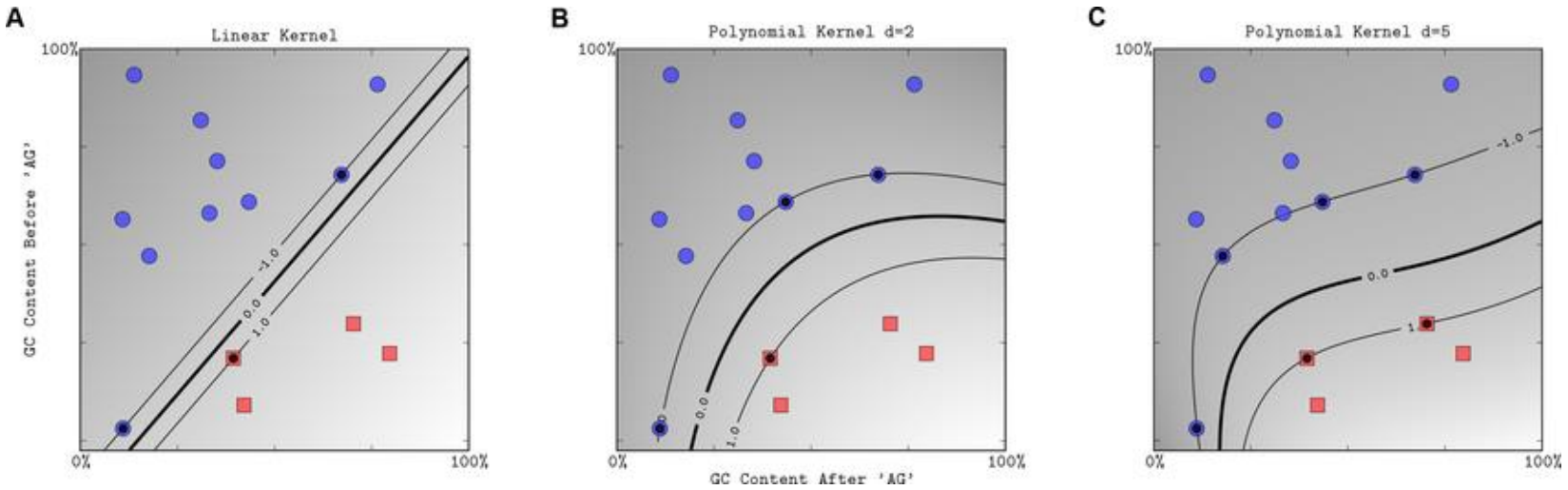


https://www.researchgate.net/figure/Figure-C-Diagram-of-2D-Hyperplane-vs-3D-Dataset-6_fig1_345314400

Kernels

1. Linear kernel
2. Polynomial Kernel

$$K(x_1, x_2) = (x_1^T x_2 + c)^d$$

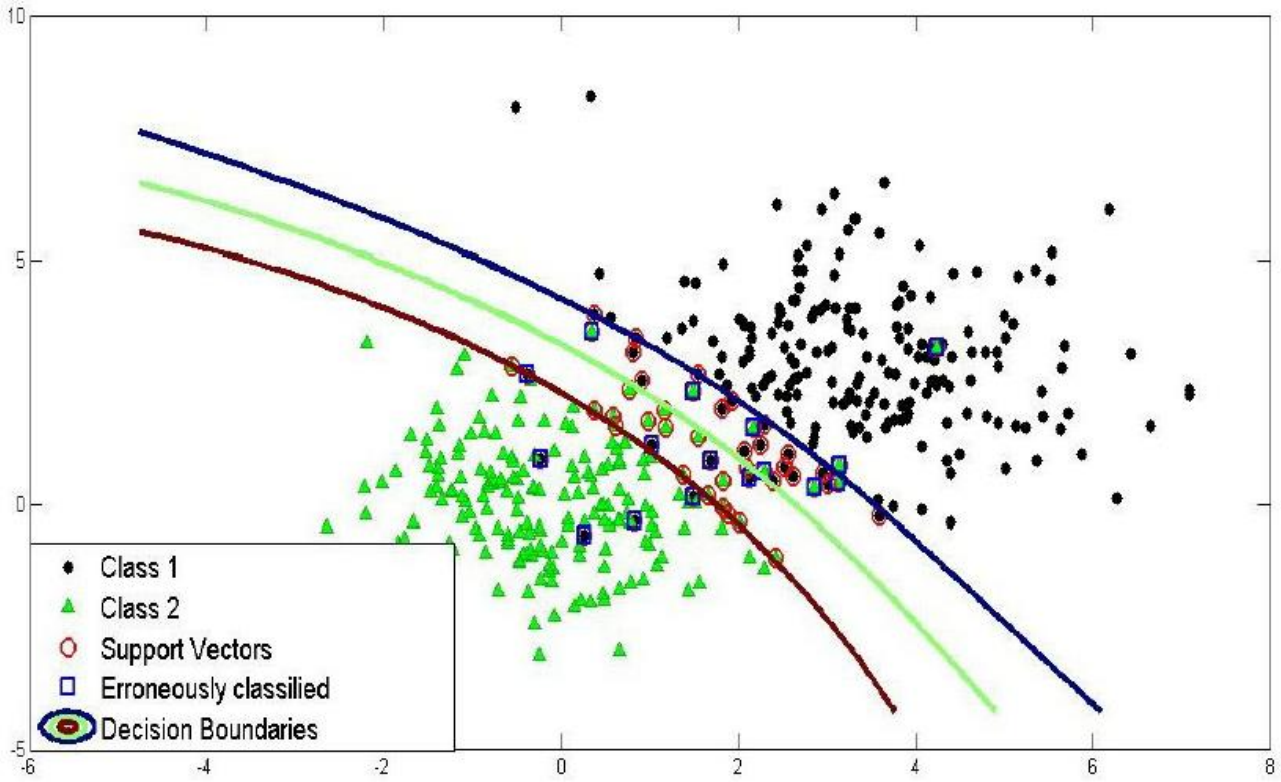


https://www.kaggle.com/code/prashant111/svm-classifier-tutorial

Kernels

3. Radial Basis Function Kernel

$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$



https://www.kaggle.com/code/prashant111/svm-classifier-tutorial

Pros & Cons

Pros

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

Cons

- Isn’t suited to larger datasets as the training time with SVMs can be high
- Less effective on noisier datasets with overlapping classes