Data-Driven Problem Solving in Mechanical Engineering

Support Vector Machine

Masoud Masoumi

ME 364 - Spring 2022

Department of Mechanical Engineering
The Cooper Union for the Advancement of Science and Art

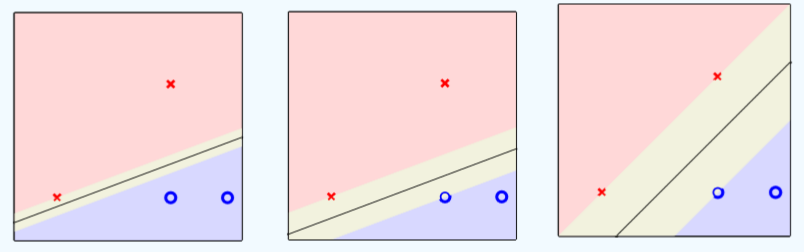
April, 2022

SVM

A classifier basically separates different classes in the data using **decision boundaries** and by carving feature space into regions, so that all the points within any given region are destined to be assigned the same label.

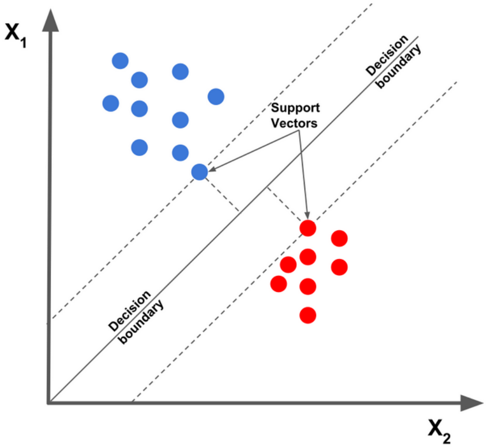
Support Vector Machine (SVM) is a supervised learning algorithm mostly used for classification.

In **SVM**, we seek to maximize the margin for the separator between the two classes.



SVM

The channel between two classes is defined by a small number of data points as opposed to logistic regression, where all the points contribute to best position of the line. These contact points are the **support vectors** defining the channels.



ref: https://www.learnopencv.com/support-vector-machines-svm

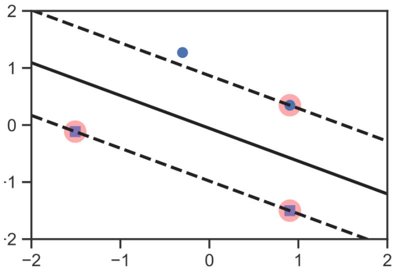
SVM

If we have 1D data, we would separate the data using a **single threshold value**. If we have 2D data, we would separate the data using best **line**. If we have 3D data, the output of SVM is a **plane** that separates the two classes. Finally, if the data is more than three dimensions, the decision boundary is a **hyperplane**.

SVM solves an optimization problem such that

- Support vectors have the greatest possible distance from the decision boundary (i.e. separating hyperplane)

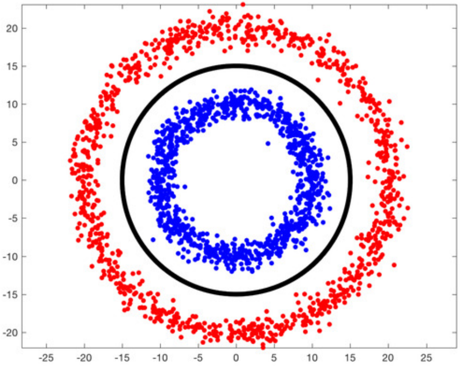
- The two classes lie on different sides of the hyperplane



If you are interested in knowing more about how the process works, I encourage you to look at this link https://www.jeremyjordan.me/support-vector-machines/ or watch this great lecture https://www.youtube.com/watch?v=_PwhiWxHK8o

Kernel Trick

What if the data is not separable by a hyperplane? For example, in the following figure, two classes represented by the red and blue dots are not linearly separable. The decision boundary shown in black is actually circular.



Kernel Trick

In such a case, we add a new dimension using the **Kernel Trick** (see here: https://en.wikipedia.org/wiki/Kernel_method) where we add a new dimension to existing data and if we are lucky, in the new space, the data is linearly separable.

For example for our case, we can add a new dimension using $z = e^{-\gamma(x^2+y^2)}$

The parameter γ controls the amount of stretching in the z direction. See an animation here

The function we used for adding a new dimension to the data is **Gaussian Radial Basis Function** or a **Radial Basis Function with a Gaussian kernel**. Some here for other popular kernels typically used