SUPPORT VECTOR MACHINES

Authors: G. H. Zerwes, *Universidade Federal do Espirito Santo, Vitória, Brazil*

Research Supervisor: L. Homri, *Arts et Métiers ParisTech, Metz, France*

1 INTRODUCTION

The Support Vector Machine (SVM) is an algorithm used best to classify data into two categories. It works by defining a decision boundary between the data points belonging to each class. This decision boundary should have the largest possible margin, that is to say, the boundary should be the furthest away possible from any other points.

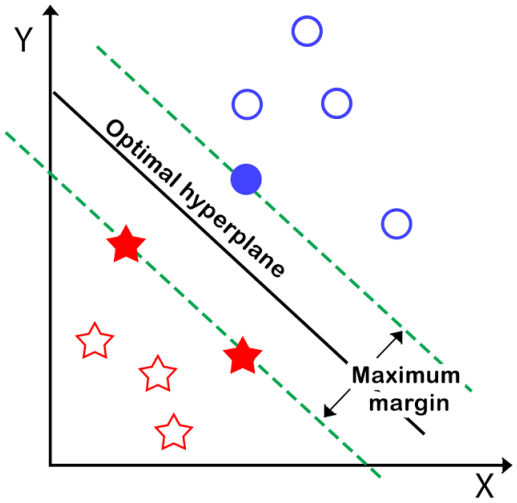

Figure 1: Example of decision boundary

An example of this can be seen in figure 1. The different classes are represented by the shapes (red star and blue circle), and the margin is highlighted as the distance between the hyperplane (decision boundary) and the nearest point. The optimal hyperplane is achieved by maximizing this margin. Another important concept is the support vector, which is the data points that are closest to the margin. In the example figure, they are represented by the filled shapes.

At last, it's important to be aware of the function of the kernel. A kernel is simply a function that transforms the data into a space so that finding the boundary is an easier task to perform. There are different types of kernels, and usually, the quality of the model fitting is heavily dependent on the kernel choice.

2 Pseudo-code

To use this algorithm on your data, the pseudo-code below shows a procedure that should be followed.

```
#Step 1: Gather the data. The X data is the independent variable and should be already
    pre-treated. The Y array is the vector containing the labels for each class.

X = [x_value_1, x_value_2, ...]
Y = [y_value_1, y_value_2, ...]

#Step 2: Fit the model.
model = SVM.fit(X,Y)

#Step 3: Make new predictions.
y_prediction = model.predict(X)
```

3 Other recommendations

This algorithm's computational cost scales up fast with the number of samples, so it's not recommended to use it when there are more than a couple thousand samples.

The kernel and regulation parameters choice have big impacts on the performance of the model. The C value of 1 is usually a good starting point, however, it should be decreased if there are large amounts of noise present. The choice of a kernel depends on the type of data. For example, if the data is linearly separable, then a linear kernel should be preferred. Other, more complex, kernels can be used, but have higher computational costs.

The SVM algorithm is not scale insensitive, so data scaling is essential to provide good results. The standardization can be in the range of [0,1], [-1,1], or with a mean of 0 and a standard deviation of 1.

References

Bishop, C. M. (2006). *Pattern Recognition And Machine Learning*. Number 758. Springer.