

Support Vector Machine

Johnny Nguyen¹

Abstract—This electronic document permits me to synthesis the third course of data mining.

I. INTRODUCTION

This model use supervised learning, it means that we know the result. It was implemented in the sixties. The goal is to be better than the perceptron algorithm. It can be use with very high dimension space.

II. PRINCIPLE

The goal is to find the best vector between two cluster that separate properly the points while maximizing the margin. Support Vector Machine contains a unique solution. In Python, we can use library sklearn to use this algorithm. The noise is ignored by this model. The kernel trick consists in add dimension to separate with a line the two clusters. We can use our own kernel to do this trick.

III. SIMPLE BINNING METHOD

The goal is to train the SVM classifier with a **training set**. Then, we apply a **validation set** to compute the distance of each sample. Finally, we use an histogram which show us that our classifier works. The Platt's method consists in use a sigmoid filter to have better result.

IV. CONCLUSION

This method doesn't need a huge dataset to be efficient.

ACKNOWLEDGMENT

Thanks to Diane Lingrand for his work.

REFERENCES

- [1] <https://moodle.polytech.unice.fr/course/view.php?id=28>