

# Optimisation with Evolutionary Algorithms in Data Mining

Johnny Nguyen<sup>1</sup>

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

## I. INTRODUCTION

L'application de méthode complexe pour trouver la presque meilleur solution. Il faudra forcément maximiser ou minimiser quelque chose. L'utilisation d'heuristique pour trouver une méthode de résolution à tout prix.

Un problème d'optimisation se fabrique de cette façon :

- Entrée : plusieurs solutions,
- Sortie : trouver la meilleure solution,
- Une fonction  $f$  est déterminée, cette fonction peut-être complexe (multi-objective).

La méthode d'optimisation peut-être exact, c'est à dire prendre plusieurs années. Et, elle peut aussi être une méta-heuristics, c'est à dire, trouver un ou plusieurs résultats en un temps limité.

## II. EVOLUTION ALGORITHMS

Utiliser la théorie de Darwin correspondant à la reproduction et la mutation d'un génome.

- Le choix des parents,
- La production d'un nouveau génome en ajoutant les deux génomes.

Ces deux étapes sont faites en boucle pour établir une génération. Une seule génération peut prendre plusieurs années à être effectué, c'est pourquoi nous devons choisir un bon framework.

### A. L'exemple du jouet

Nous devons l'établir sur une feuille. Nous mettons les données dans un graphique et nous extrayons la forme issue de ces points.

## III. 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.

## IV. 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.

## V. 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>

<sup>1</sup>This work was not supported by any organization