1 november 2021

Team Emo:

Julian Groen
Zahir Bholai
Koen de Bruijn
Jaap van Gestel
Breno van Tricht
Yuri Lamijo

Learning Lab

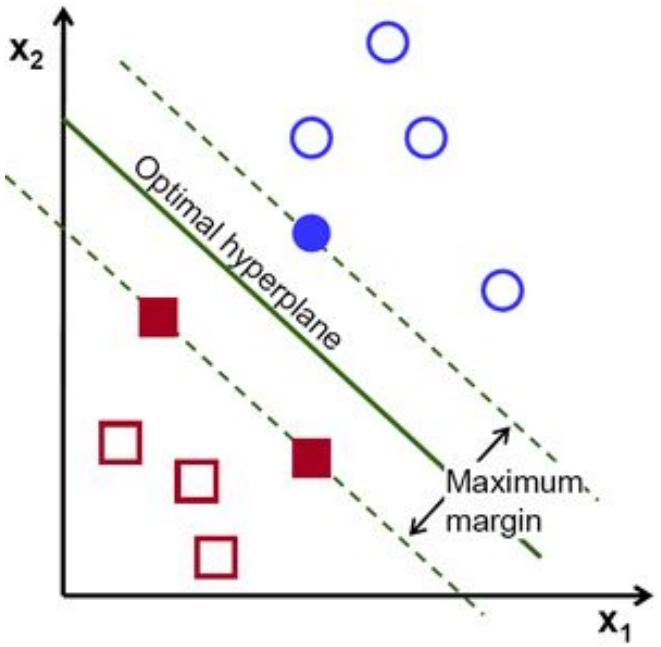
Support Vector Machines (SVM)

DE HAAGSE HOGESCHOOL

Table of contents

Intro

- SVM definition



DE HAAGSE HOGESCHOOL

Advantages SVM

1. Works relatively well when there is a clear margin of separation between classes
2. More effective in high dimensional spaces
3. Effective in cases where the number of dimensions is greater than the number of samples
4. Relatively memory efficient

Disadvantages SVM

1. Not suitable for large datasets
2. Does not perform very well when the dataset has more noise i.e. target classes are overlapping
3. It will underperform when the number of features for each data point exceeds the number of training data samples

Hyperparameters

- Kernel
- C
- Gamma

Classification - linearly separable data

Weight determines obesity

We can already see where to separate the classes.
No ML necessary.

Divide the difference between the outermost blue observation and the innermost red one.

Maximizing the margin between the threshold and the data is done using a **maximum margin classifier**

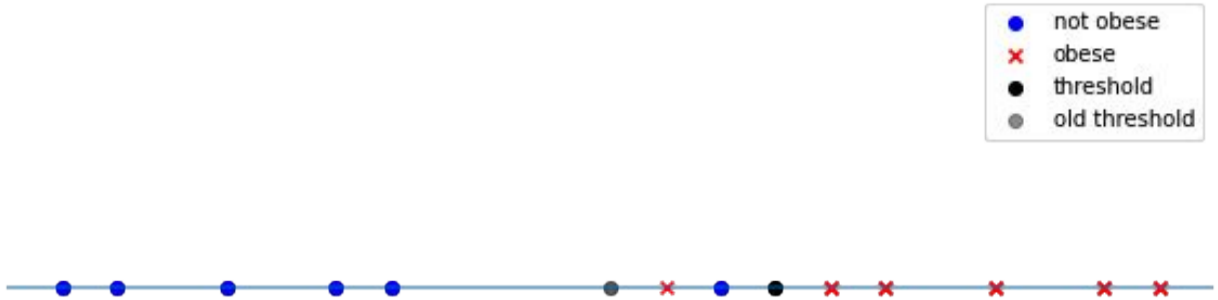
DE HAAGSE HOGESCHOOL

Classification - linearly separable data

When we move our threshold the margin becomes small.

When we get a new observation...

Problem: outliers



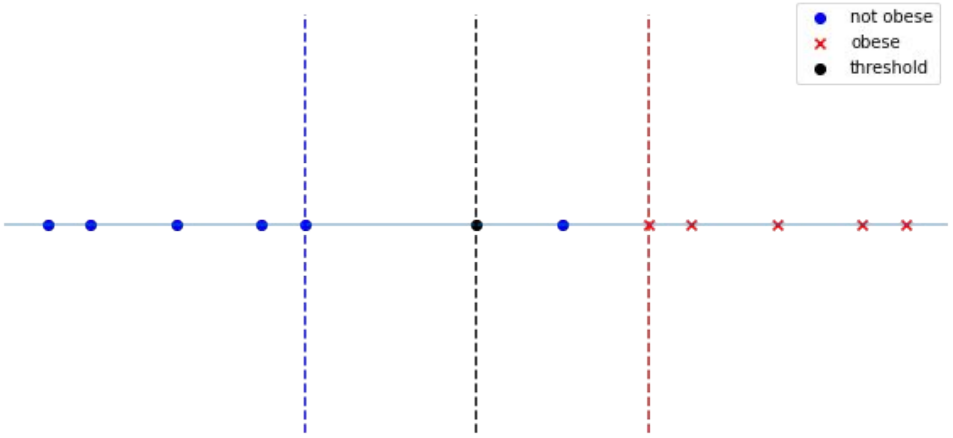
Our old threshold would correctly classify it as obese.

Our new threshold does not.

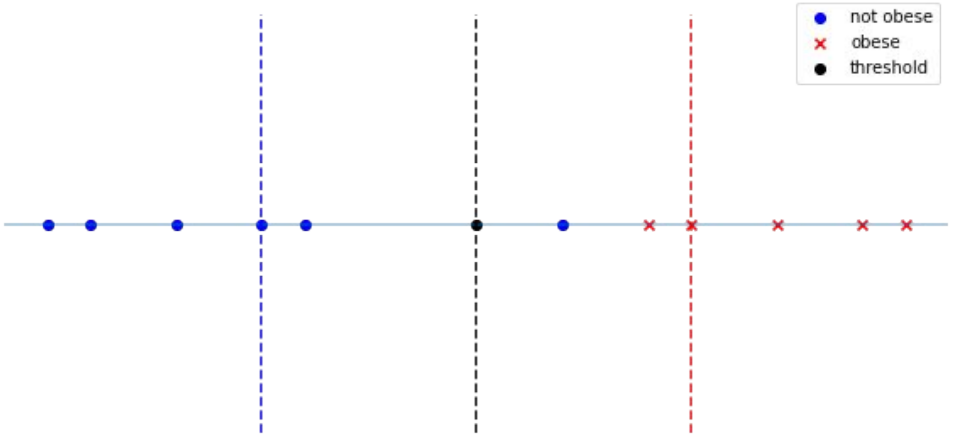
DE HAAGSE HOGESCHOOL

Classification - linearly separable data

- Solution: allow for misclassification using a **soft margin classifier**



But how do we determine which observations might be misclassifications and how do we determine the soft margin?



DE HAAGSE HOGESCHOOL

Classification - linearly separable data

- Answer: **Cross-validation**

Determine how many misclassifications are allowed and how many observations are in the soft margin.

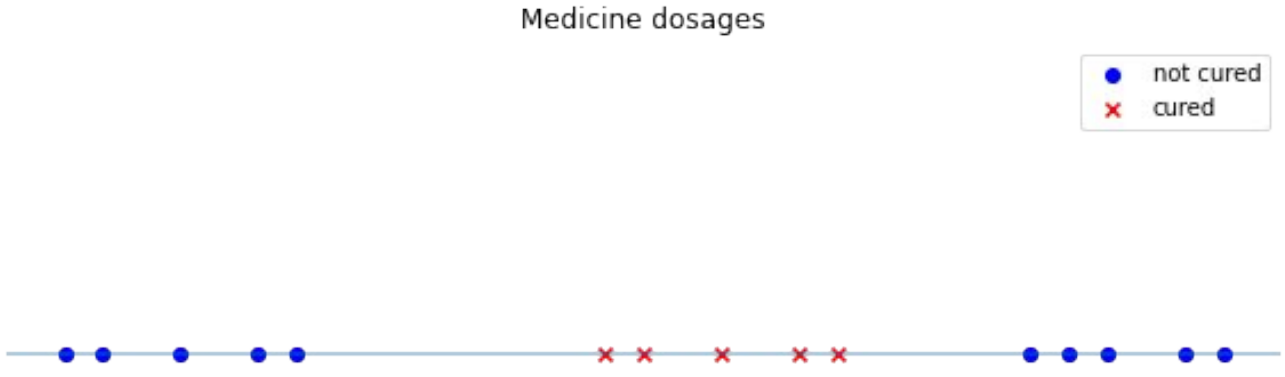
Using a soft margin to determine the location of the threshold is also known as: using a **Support Vector Classifier**

The support vectors are the observations outside the soft margin that impact the threshold the most.

Classification - non-linearly separable data

Sick people can get cured with the correct dosage of medicine.

Can we still place a threshold?



DE HAAGSE HOGESCHOOL

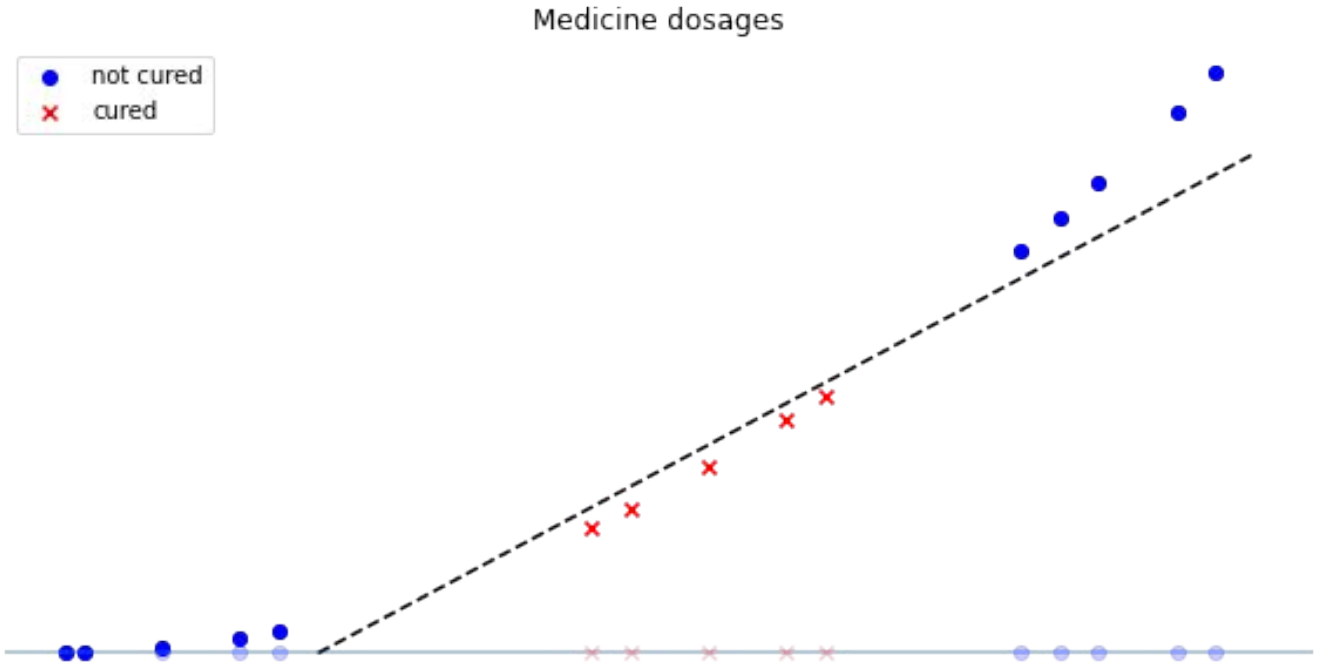
Classification - non-linearly separable data

Enter: **Support Vector Machines**

Kernel Trick:

1) Move to a higher dimension
2) Transform the data with a function

The function depends on the kernel



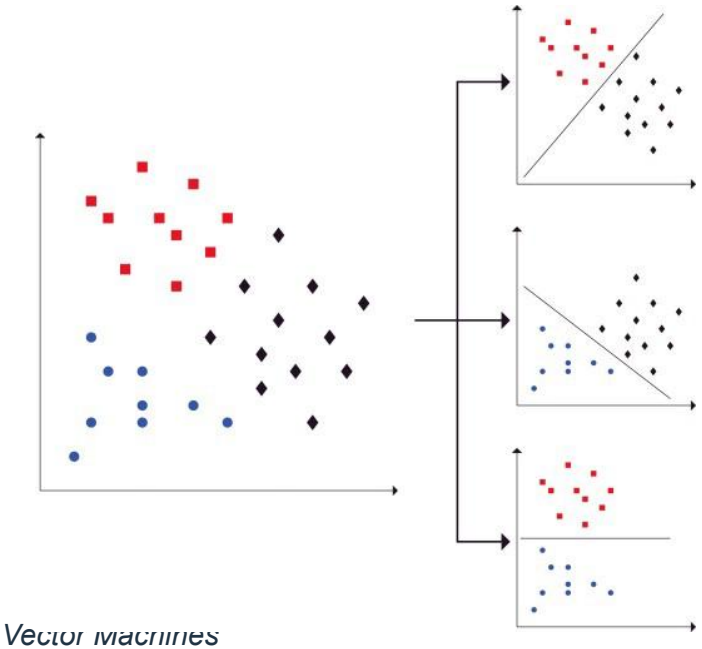
DE HAAGSE HOGESCHOOL

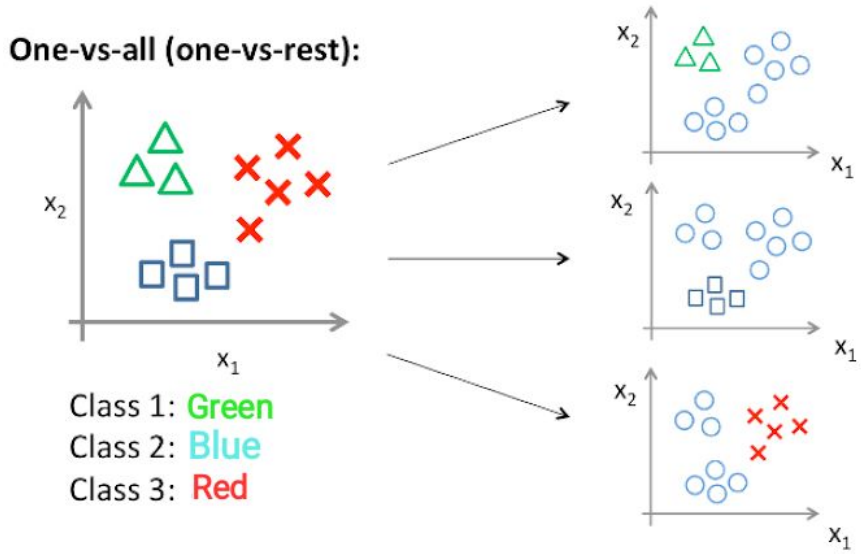
Classification - Kernel

- Kernel functions systematically find Support Vector Classifiers in higher dimensions
- There are many different kernels:
  - *Polynomial Kernel,*
  - *Gaussian Kernel*
  - *Radial Basic Function (RBF)*
  - …
- We won't cover them here but:
- **Choosing the right kernel is important**

Classification - Multiclass classification

- SVMs break down a multiclass classification problem into multiple binary classification problems
- Can either be done through *one-to-one OR one-to-rest* approach
-





HOGESCHOOL

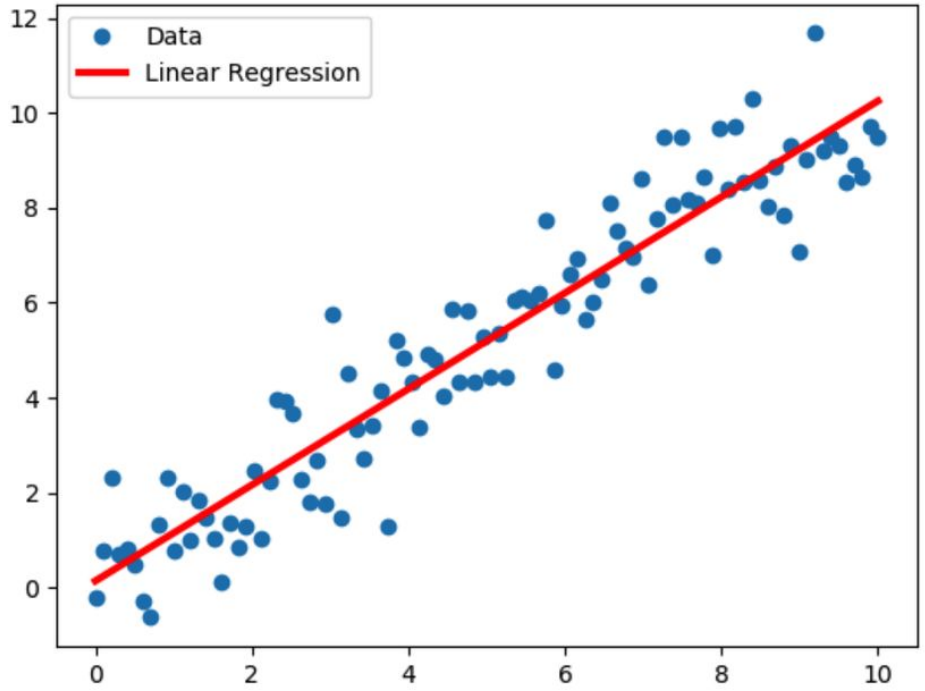
Regression - Support Vector Regression

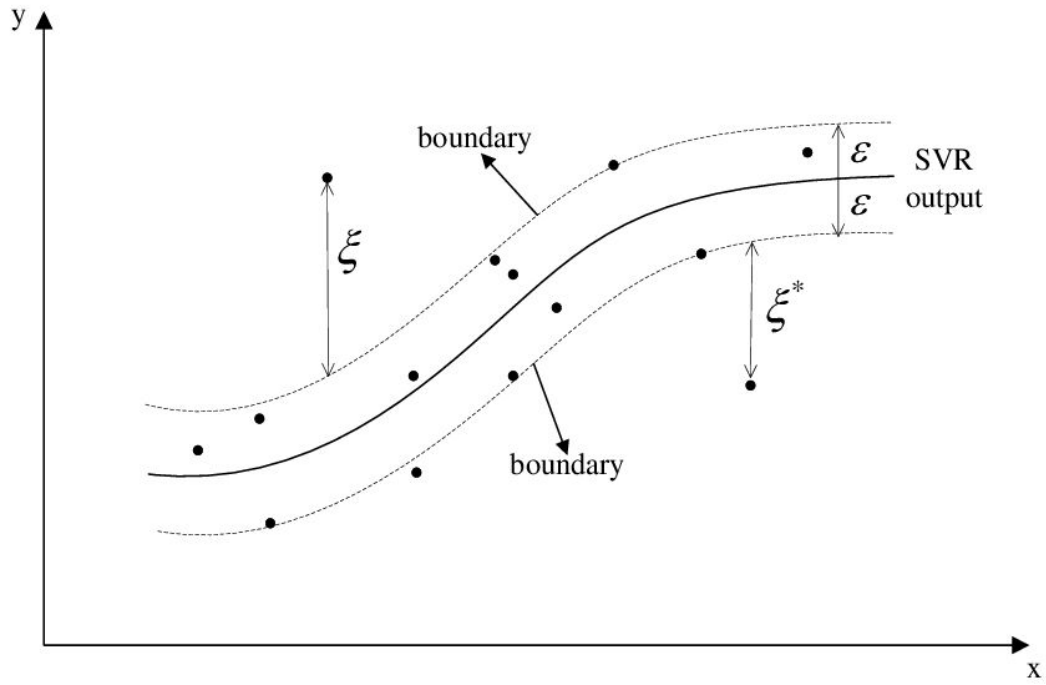
- Same principle as Support Vector Machines
- Goal SVR: find the best fit line
- Best fit line: hyperplane that has the maximum number of points
- Key hyperparameters SVR: Hyperplane, Kernel and Boundary lines

Regression - Variations SVR

- epsilon-SVR
- nu-SVR
- Depending on your needs, you should choose the type of SVR that fits your machine learning problem.

Regression - Linear Regression vs SVR





DE HAAGSE HOGESCHOOL

Regression - Advantages SVR

- It is robust to outliers.
- Decision model can be easily updated.
- It has excellent generalization capability, with high prediction accuracy.
- Its implementation is easy.

Regression - Disadvantages SVR

- Like SVMs, SVRs are not suitable for large datasets.
- In cases where the number of features for each data point exceeds the number of training data samples, the SVR will underperform.
- The Decision model does not perform very well when the data set has more noise i.e. target classes are overlapping.

Notebook

- Link on Blackboard (?)

Questions?