

# CLASSIFICATION - SVM

JENS BAETENS

# WAT IS SVM?

Supervised learning

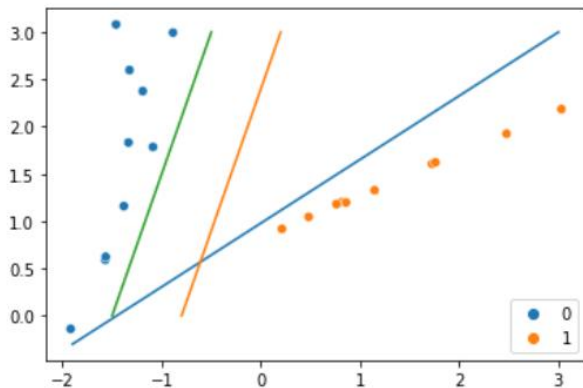
Classificatie en regressie

Classificatie door het zoeken naar hyperplanes tussen twee klassen

# PROBLEEM MET LOGISTIC REGRESSION

Kostfunctie van logistische regressie niet altijd meest robuuste oplossing

Logistische regressie ideale op de trainingsdata maar performantie op testdata is belangrijker



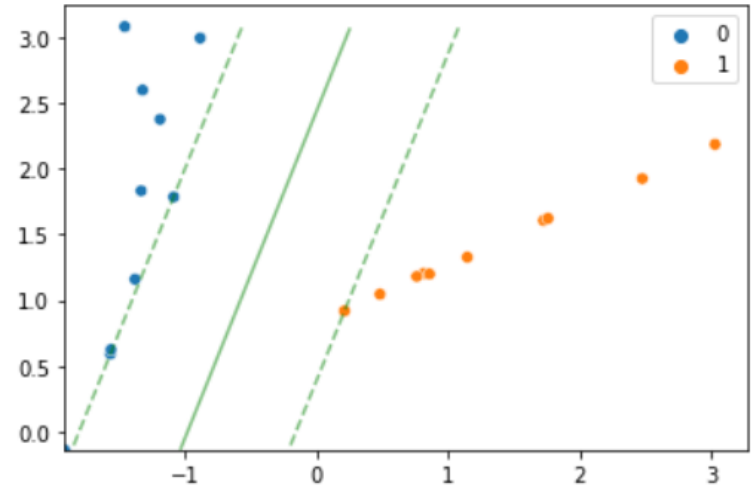
# WAT DOET SVM?

Zoekt de scheidelingslijn

Maximaliseert de marge

Enkel de dichtste punten gebruikt

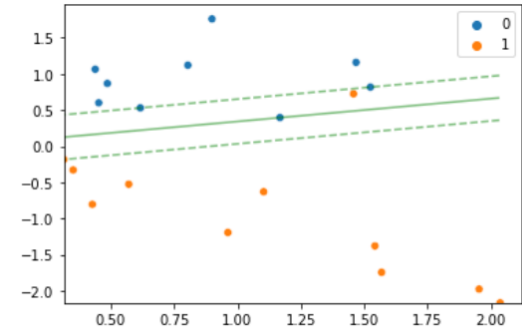
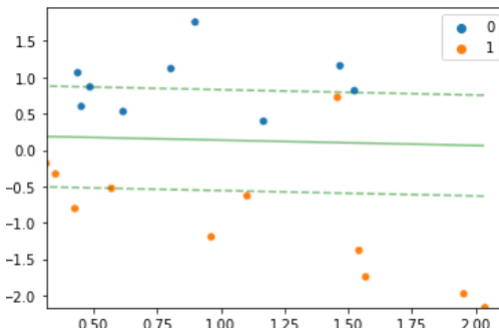
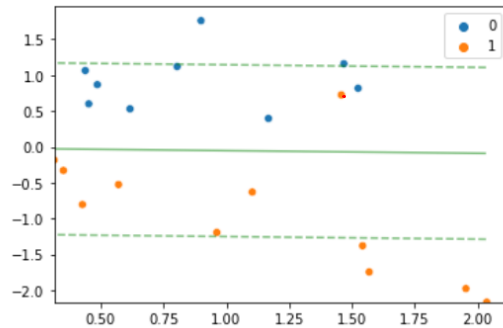
- De support vectors



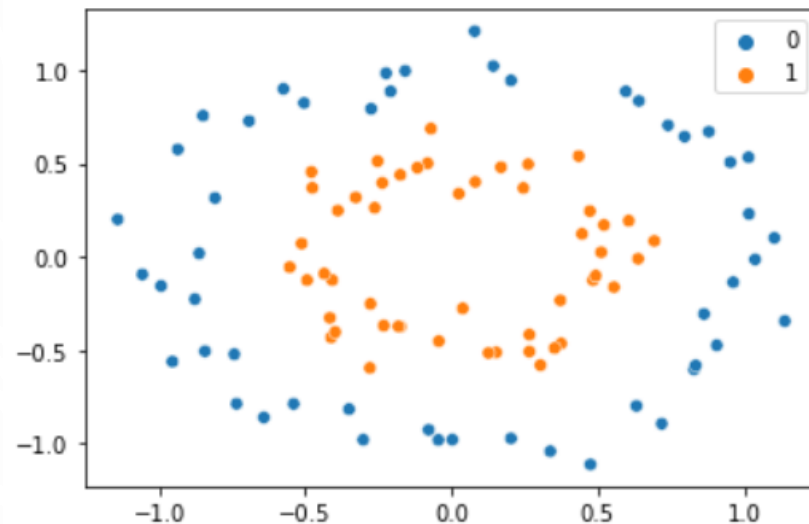
# WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

Als het “ongeveer” scheidbaar is:

- Gebruik regularisatieparameter  $C$
- Hoe groter  $C$ , hoe kleiner de toegelaten marge



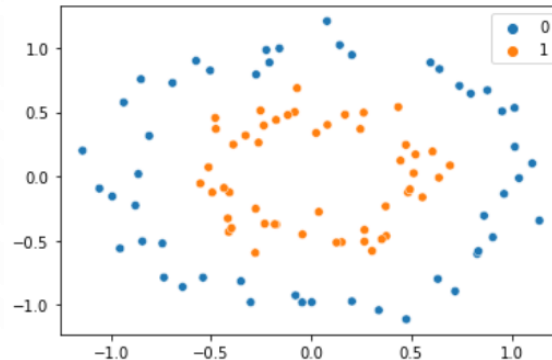
# WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS



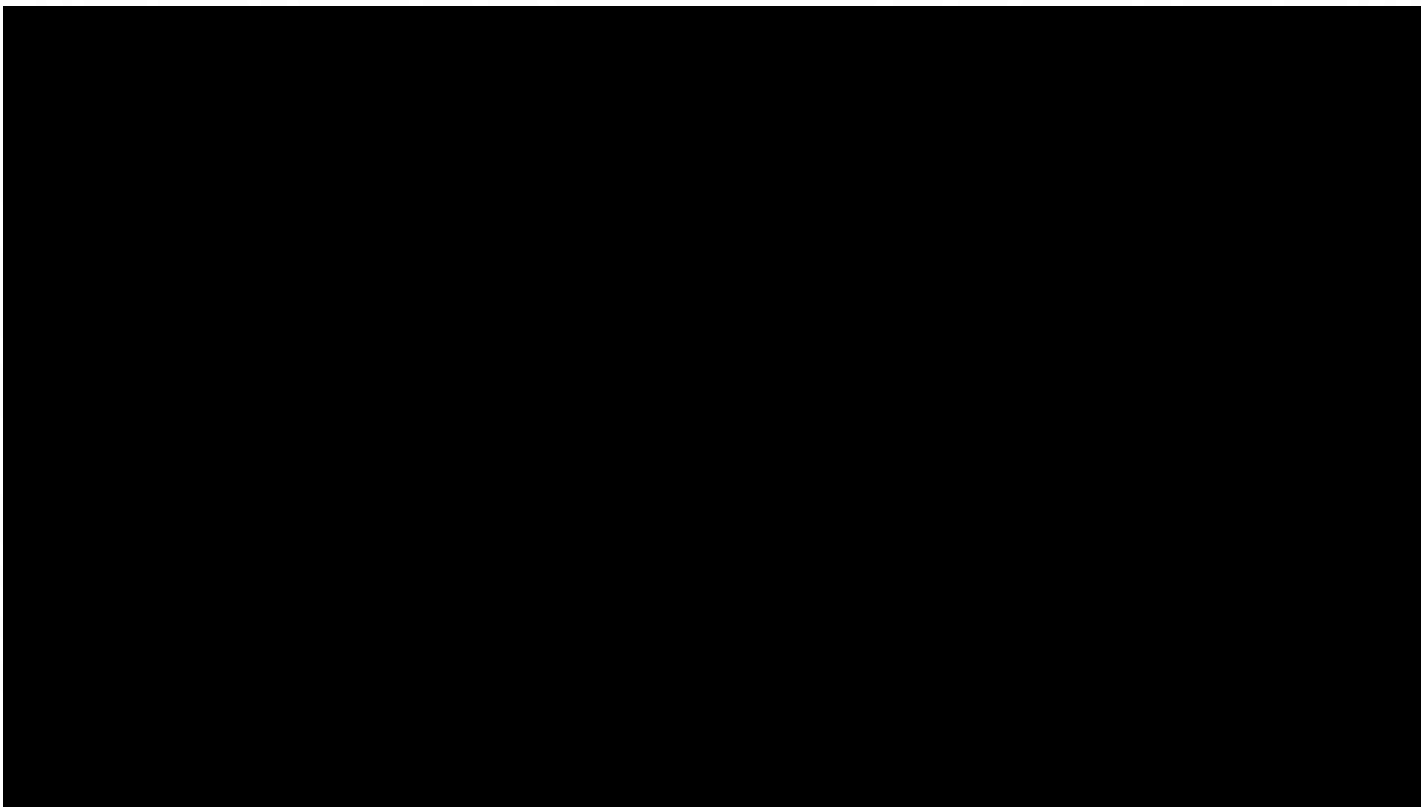
# WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

Als het helemaal niet scheidbaar is:

- Projecteer elk punt in een hogere dimensie
- De functie waarmee dit gebeurt = de kernel
- Bijvoorbeeld door de afstand tot de oorsprong te berekenen



# WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS





# WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

Verschillende kernels mogelijk:

- Lineair (geen projectie in hogere dimensie)
- Gaussiaanse kernel of Radial Basis Function (RBF): Default in sklearn
- Polynomial (vooral bij Natural Language Processing)

