

CLASSIFICATION - SVM

JENS BAETENS

Support | Machines
Vector

WAT IS SVM?

Supervised learning

Classificatie en regressie

SVC() *SVR()*

Classificatie door het zoeken naar hyperplanes tussen twee klassen

Rechte scheiding

↳ One vs One

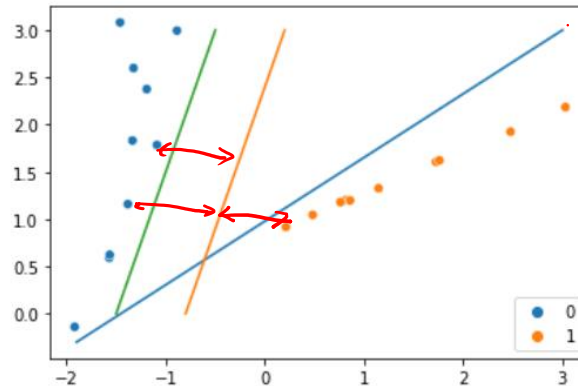
↳ One vs All

Multiclass

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



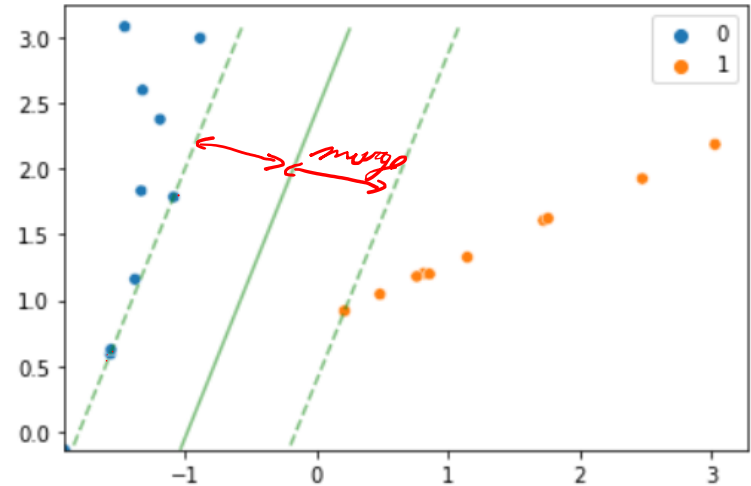
WAT DOET SVM?

Zoekt de scheidingslijn

Maximaliseert de marge

Enkel de dichtste punten gebruikt

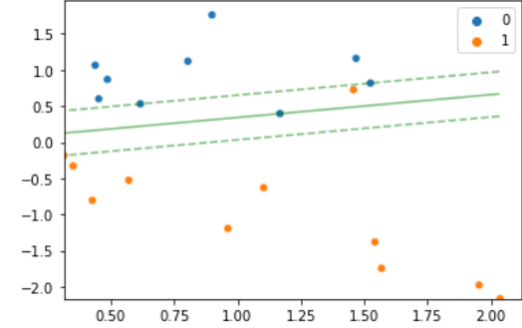
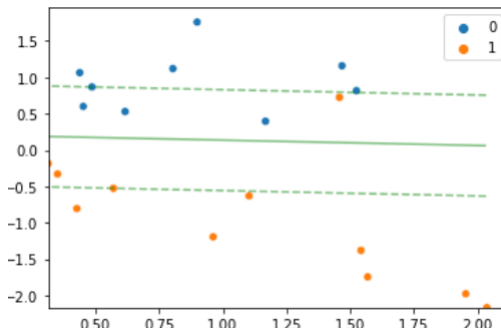
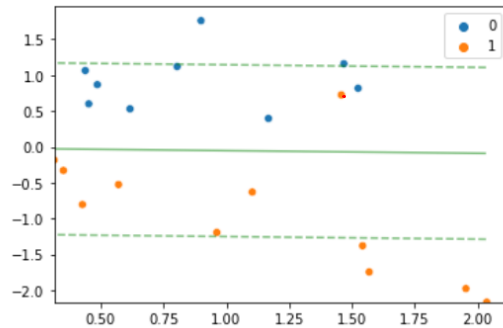
- De support vectors *genoemd*



WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

Als het “ongeveer” scheidbaar is:

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



Logistische regressie
Lineair

Sum

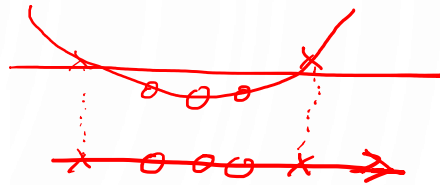
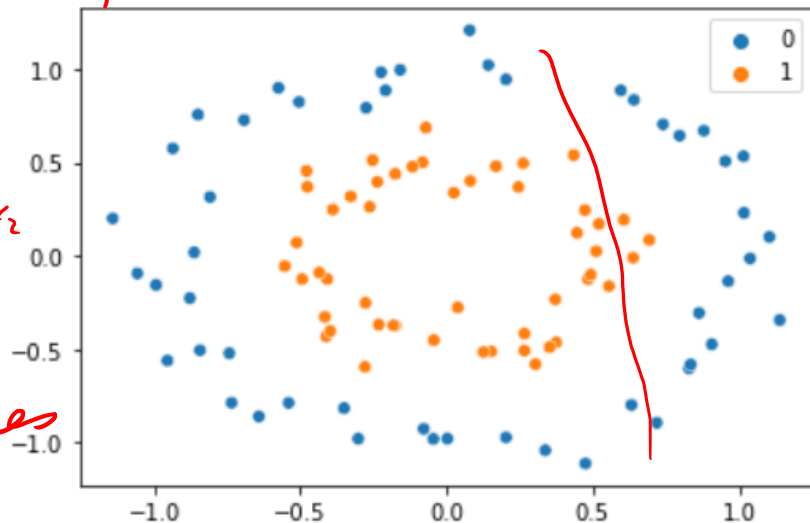
WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

→ hogere orde features | → projectie

x_1 x_2

x_1^3 x_1^2 x_1 x_2^3 x_2^2 x_2 $x_1 x_2$
 x_1^2 x_2^2 x_1 x_2 ...

Polynomial Features

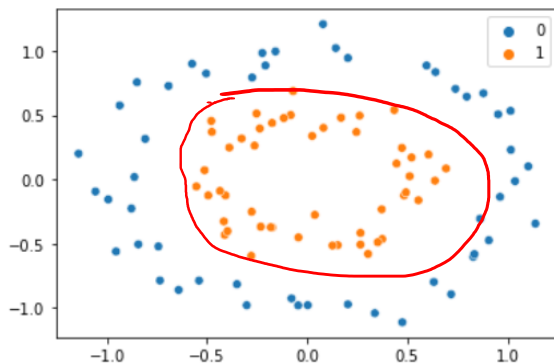


WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

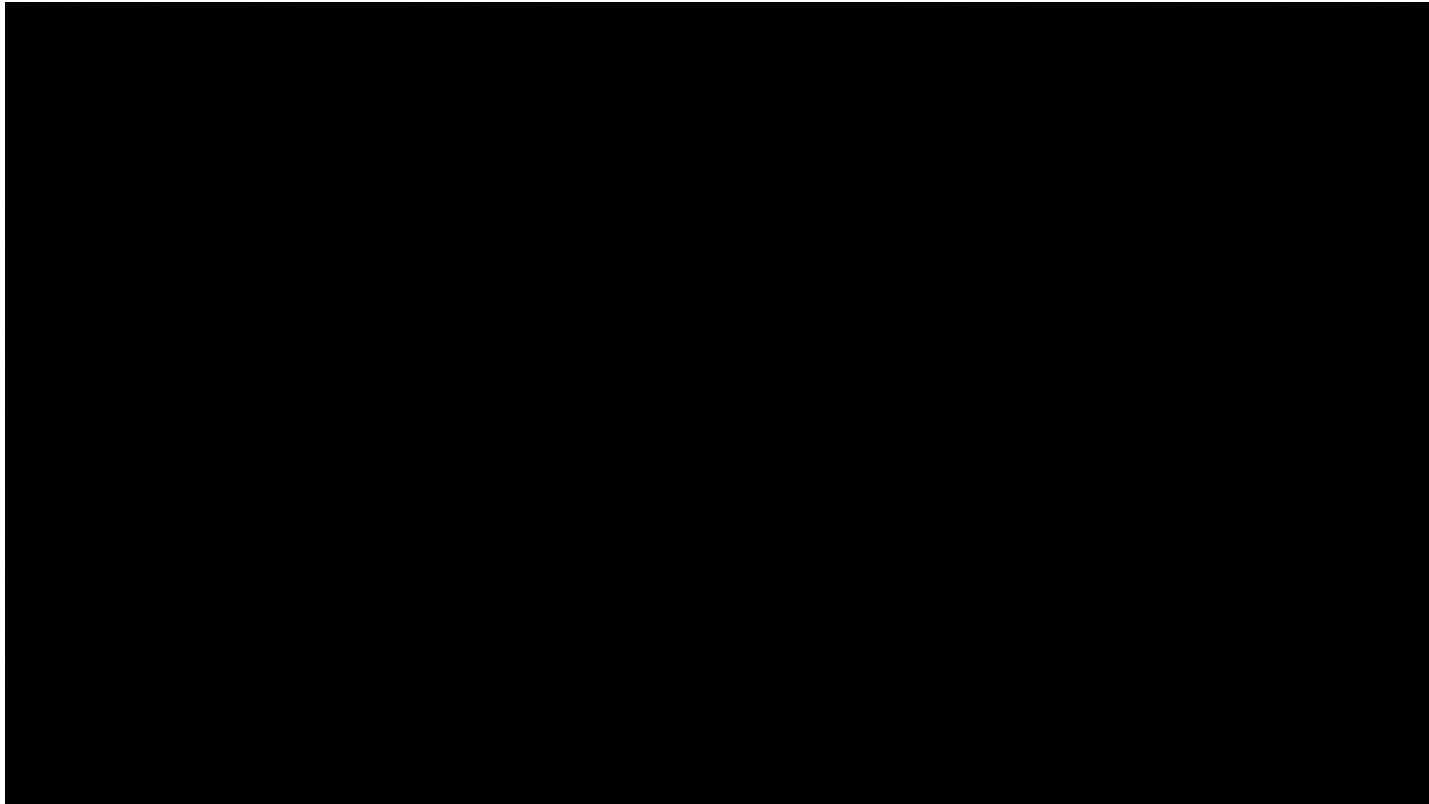
Als het helemaal niet scheidbaar is:

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

*zelf standaard
polynoom*



WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

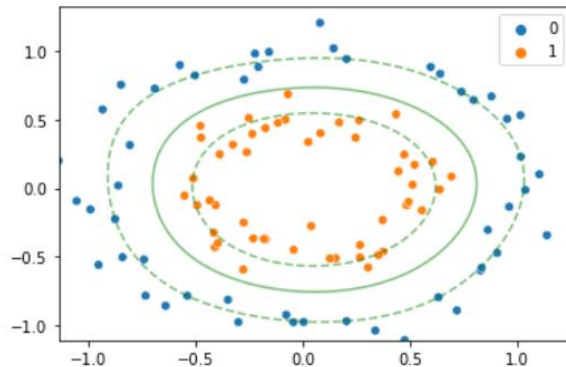


WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

→ extra hyperparameter γ → hoe "breed" is de kernel

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)



$$\gamma = 0, 1$$

underfitting

$$\gamma = 1$$

$$\gamma = 10$$

overfitting

WAT ALS DATA NIET LINEAIR SCHEIDBAAR IS

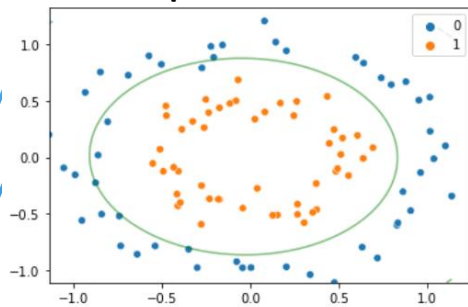
hyperparameter

Gebruik van een kernel brengt extra regularisatieparameter met zich mee

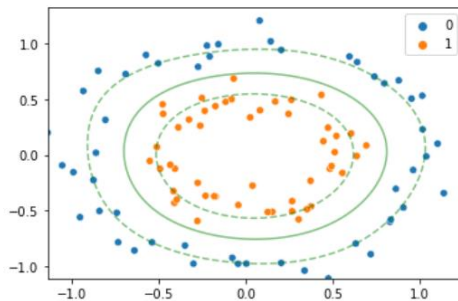
Gamma -> Breedte van de kernel / afstand waarop een punt effect heeft

- Te klein = brede kernel = Underfitting (linkse figuur)
- Te groot = smalle kernel = Overfitting (rechtse figuur)

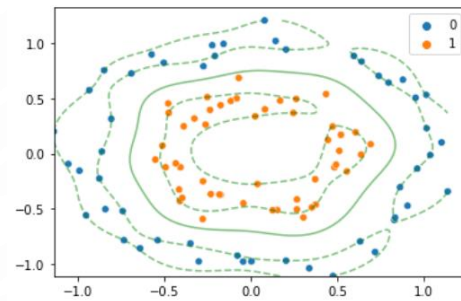
$$\gamma = 0.1$$



$$\gamma = 1$$



$$\gamma = 10$$



WAAROM SVM?

Zowel voor classificatie als regressie

Goede performantie op kleine datasets met veel features

Minder geheugen nodig door beperkt gebruik van de beschikbare data

Robuster door optimalisatie marge

↳ op de testdata

Werkt goed op niet-gestructureerde data (tekst, beelden, ...)

Minder gevoelig aan overfitting

~~test data~~

WAAROM GEEN SVM GEBRUIKEN?

Gevoelig voor outliers → *je gebruikt en bel de slachtoffer*
↳ *zijn vaak de outliers*

SVM geeft geen indicatie over de zekerheid van de classificatie

kans
Logistic Regression werkt beter op grote datasets met beperkt aantal features

→ *Explainable AI → uitleggen waarom klasse A/B*
↳ *logistic → kans*