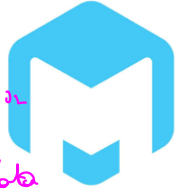


V. VAPNIK

1 → SVM

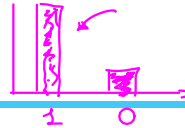
! 2 → cross-validation

3 → Imbalanced data



Machine learning

Support Vector Machines (SVM) SVC



Wouter Gevaert & Marie Dewitte

Inhoud

Support Vector Machines - SVM

Cross-validatie

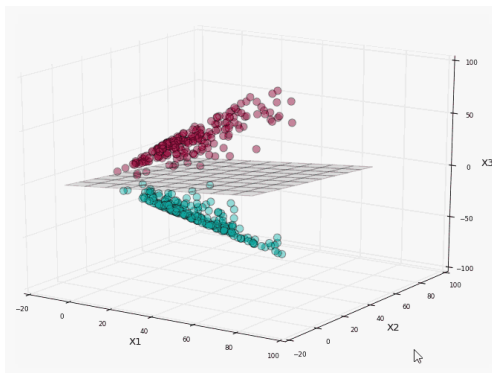
Niet-gebalanceerde data

Support Vector Machines - SVM

Support Vector Machines - SVM

Wat is een SVM?

Een SVM is een supervised ML algoritme dat zowel voor classificatie als regressie gebruikt kan worden. Classificatie gebeurt door het vinden van een hyper-plane die een optimale scheiding maakt tussen twee verschillende klassen.

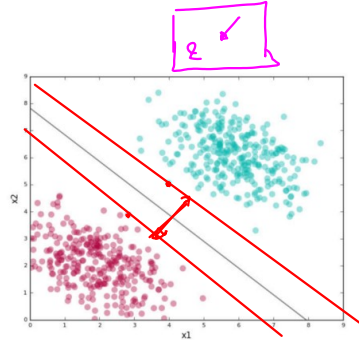
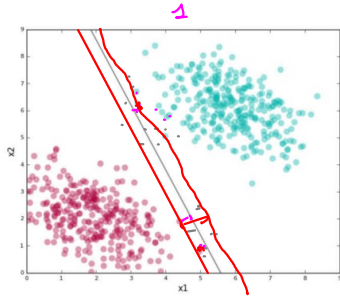


Support Vector Machines - SVM

Wat is een SVM?

→ LARGE MARGIN CLASSIFIER.

Welke classifier zou je verkiezen en waarom?



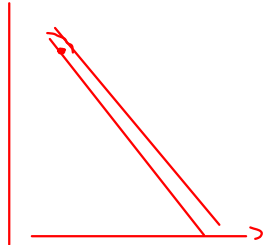
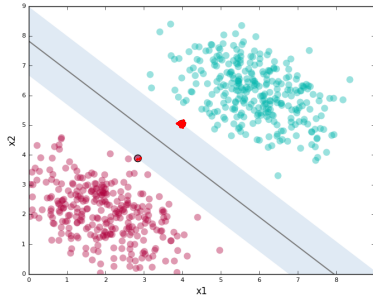
- Beide scheiden de training data perfect.
- Het gaat er niet om welke er best presteert op de training data, maar wel op de test data.
- De rechtse classifier is meer robuust.

Support Vector Machines - SVM

Hoe een SVM classificeert

- Zoek scheidingslijnen die de trainingset zo goed mogelijk scheiden.
- Kies de scheidingslijn die de grootste afstand (margin) heeft tot de punten die er het dichtst bij gelegen zijn.
- De dichtstbij gelegen punten noemen we de **support vectors**.
- SVM = Large margin classifier.

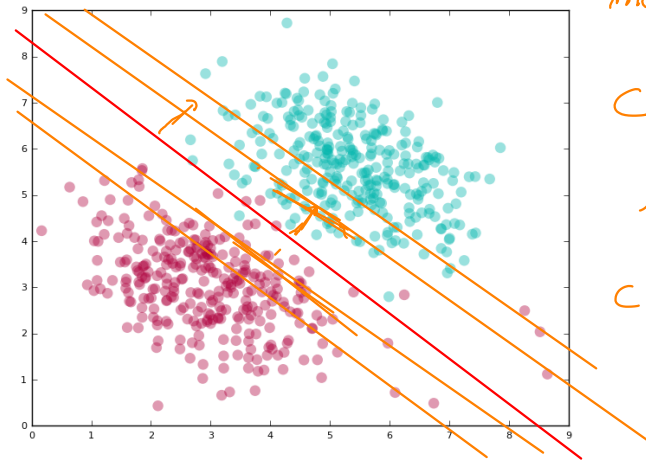
$$[x_1; x_2]$$



Support Vector Machines - SVM

Hoe een SVM classificeert

Wat als een perfecte lineaire scheiding niet mogelijk is?



C - parameter.
 $\frac{1}{C}$
 \approx breedte van de
marge.

C groot
 \Downarrow
smalle marge
 \Rightarrow overfitting

C klein
 \Downarrow
brede marge
 \Rightarrow Underfitting

Support Vector Machines - SVM

Hoe een SVM classificeert

Wat als een perfecte lineaire scheiding niet mogelijk is?

Antwoord: Werken met een **regularisatie parameter C**

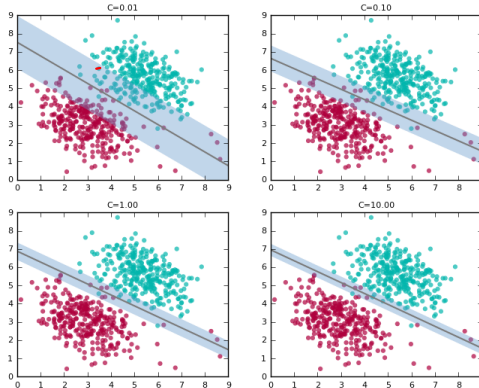
- Afweging tussen correcte classificatie op training set en een grote marge tussen de klassen (large margin).
 - Grote C-waarde: constraints zijn moeilijk te negeren \Rightarrow smalle marge.
 - Kleine C-waarde: constraints kunnen makkelijker genegeerd worden \Rightarrow brede marge.

Support Vector Machines - SVM

Hoe een SVM classificeert

Wat als een perfecte lineaire scheiding niet mogelijk is?

Antwoord: Werken met een **regularisatie parameter C**



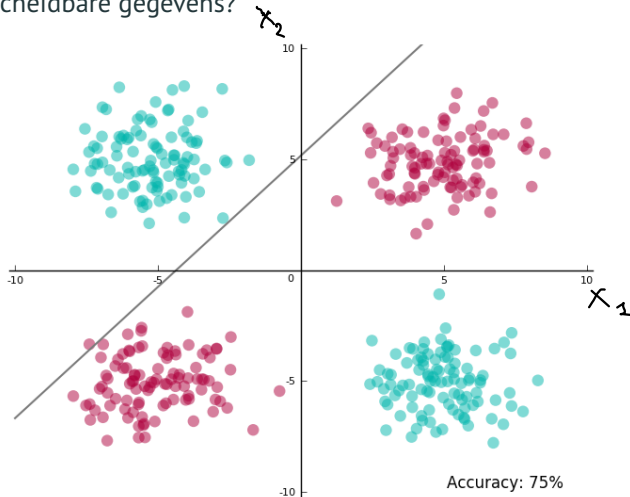
Support Vector Machines - SVM

Hoe een SVM classificeert

XOR - probleem

KERNELS

Wat bij niet-lineair scheidbare gegevens?



Support Vector Machines - SVM

Hoe een SVM classificeert

Wat bij niet-lineair scheidbare gegevens?

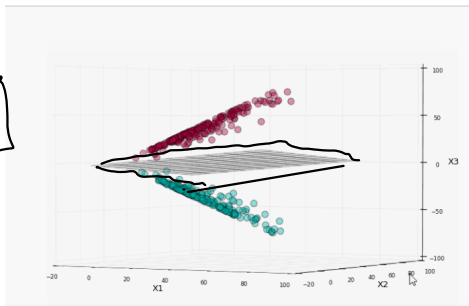
Oplossing: transformeer de data naar een hogere dimensie gevolgd door lineaire scheiding.

Bijvoorbeeld:

$$\left\{ \begin{array}{l} X_1 = x_1^2 \\ X_2 = x_2^2 \\ X_3 = \sqrt{2}x_1x_2 \end{array} \right\} \text{ KERNELS}$$

$$+x_1 + x_2 = +$$

$$+x_1 - x_2 = -$$

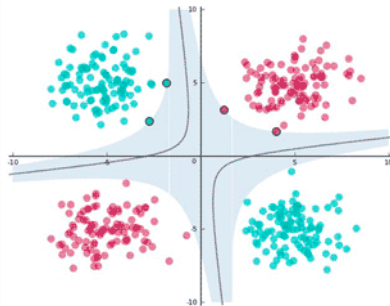
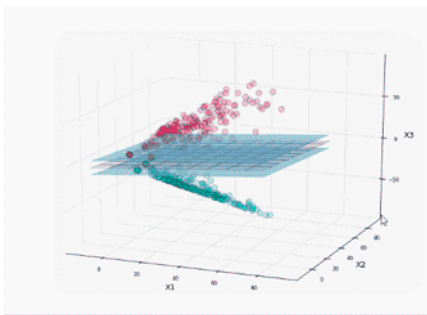


Support Vector Machines - SVM

Hoe een SVM classificeert

Wat bij niet-lineair scheidbare gegevens?

Oplossing: transformeer de data naar een hogere dimensie gevolgd door lineaire scheiding.



Support Vector Machines - SVM

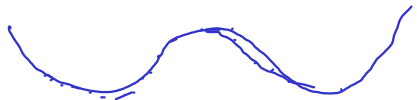
Kernels

Projecteer de data in een hogere dimensie en probeer lineair te scheiden.

Meest gebruikte kernels:

- • RBF - Radial Basis Function (Gaussiaanse kernel)
- • Polynomial kernel
- • Histogram kernel
- • Lineaire kernel = SVM zonder kernel

→ Default bij SVC

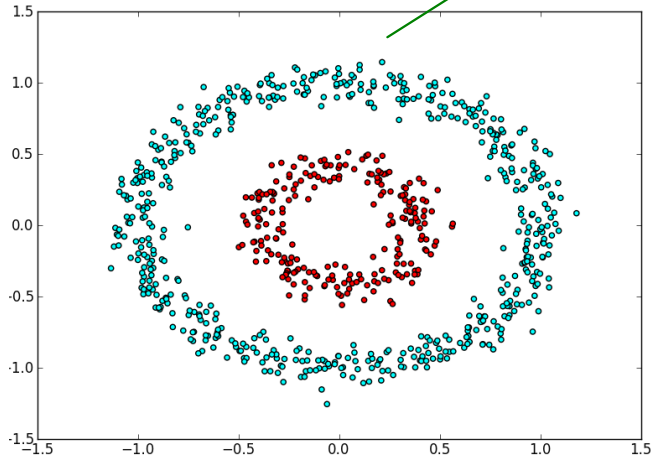


Support Vector Machines - SVM

Kernels

DONUT PROBLEM

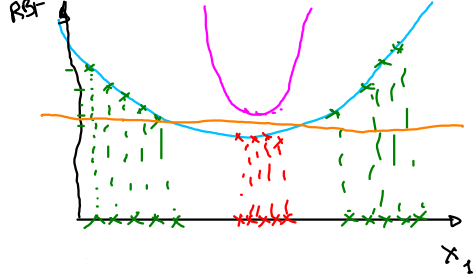
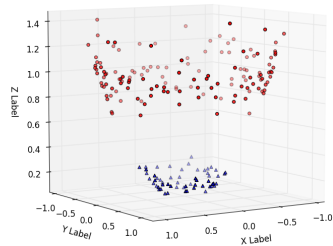
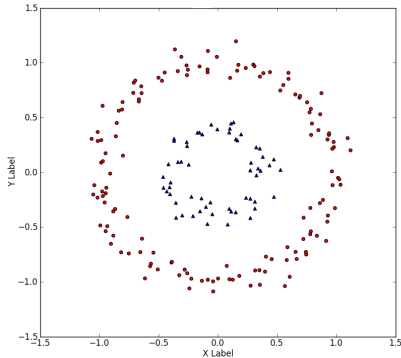
Voorbeeld: Hoe kan je onderstaande klassen scheiden?



Support Vector Machines - SVM

Kernels

Voorbeeld: Gebruik van een Gaussiaanse (RBF) kernel



Support Vector Machines - SVM

Kernels

Parameter gamma regelt de breedte van de RBF kernels

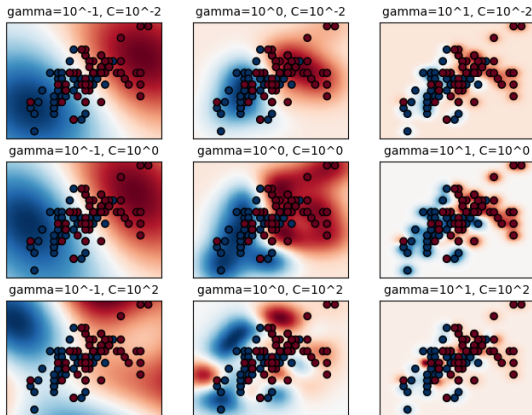
- Kleine gamma \Rightarrow brede RBF kernels. Te kleine gamma leidt ertoe dat het model de complexiteit van het model niet kan capteren (underfitting).
- Grote gamma \Rightarrow smalle RBF kernels. Te grote gamma leidt tot overfitting.

Bij gebruik van een RBF kernel: **feature scaling** (= normalisatie) toepassen.

Support Vector Machines - SVM

Kernels

Parameter gamma regelt de breedte van de RBF kernels



Support Vector Machines - SVM

Hyperparameters

Implementeren van een SVM:

- Test lineaire kernel (geen kernel) en RBF kernel.
- Tune de parameter C.
- Bij gebruik van RBF kernel: tune zowel de parameters C als gamma.

Support Vector Machines - SVM

SCALING

Motivatie voor het gebruik van SVM

- Kan zowel gebruikt worden voor regressie als classificatie (en zelfs clustering).
- Werkt goed op kleine datasets (in tegenstelling tot neurale netwerken en deep learning).
- Is nog altijd effectief wanneer het aantal features groter is dan het aantal training samples.
- Het werkt goed bij een groot aantal features (high dimensional space).
- Gebruikt niet alle training examples tijdens training \Rightarrow geheugenefficiënt.
- Geen lokale minima/optima, maar globaal optimum.

Logistic regression vs. SVM

Wanneer welke classifier kiezen?

- Wanneer het aantal features groot is ten opzichte van het aantal training samples: gebruik logistic regression of SVM zonder kernel (= lineaire kernel).
- Wanneer het aantal features klein is en het aantal training samples behoorlijk: gebruik SVM met RBF kernel.
- Bij een klein aantal features met een groot aantal training samples: creëer meer features en gebruik logistic regression of SVM zonder kernel (= lineaire kernel).

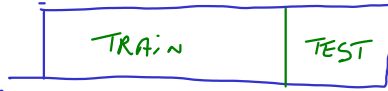
- ⇒ Hyperparameter tuning
- Evaluat. van modellen met weinig data.

Cross-validatie

Cross-validatie

Verskillende types cross-validatie

- Hold Out

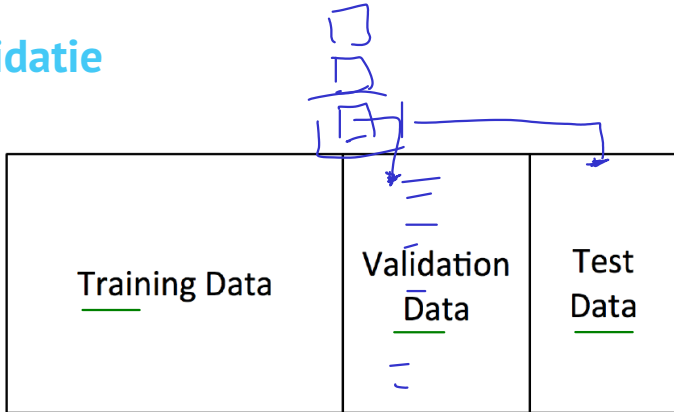


- K-fold cross-validation

- Leave One Out cross-validation
- Bootstrap cross-validation

Cross-validatie

Hold out



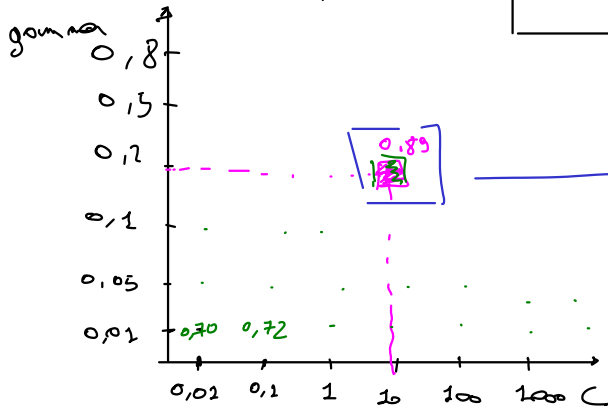
- Training data: om model mee te trainen.
- Validation data: tuning van hyper parameters en model selection.
- Test data: uiteindelijke test van het best gevalideerde model op nog nooit geziene data.

SVC (RBF, C, γ)

$C = ?$

$\gamma = ?$

GRID-SEARCH

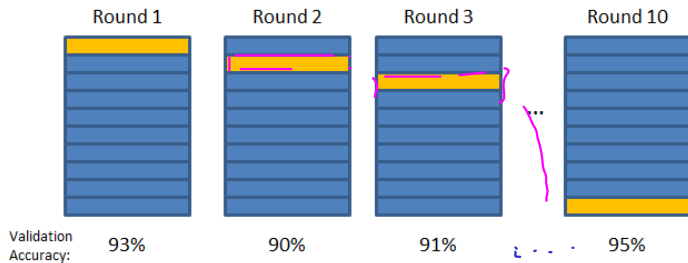
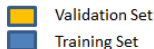


Model selection

Cross-validatie

K-fold cross validation

- Voorbeeld: 10-fold cross-validation

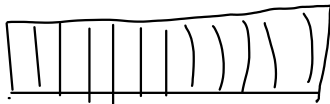


Final Accuracy = Average(Round 1, Round 2, ...)

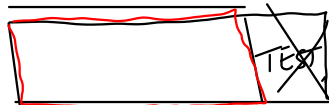
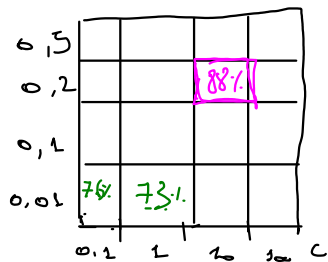
$= \text{gemiddelde} = 93\%$

$K = 10$

$K = 5 \rightarrow k = 20$

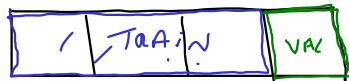


SVC (RBF ; $C = 0.1$; $\gamma = 0.01$)



$K = 4$

Ronde 1



ALL

80%

Ronde 2



70%

Ronde 3



73%

Ronde 4



80%

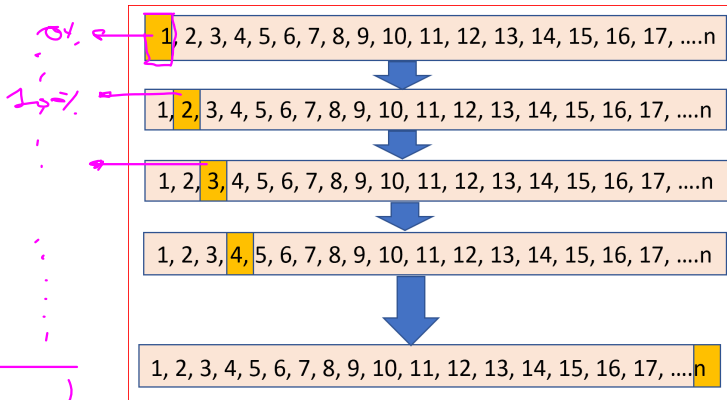
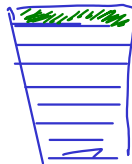
gen:ALL

76%

Cross-validatie

Leave one out cross-validation

- $K = N$ (aantal examples in de training set)
- Evenveel rondes als aantal examples

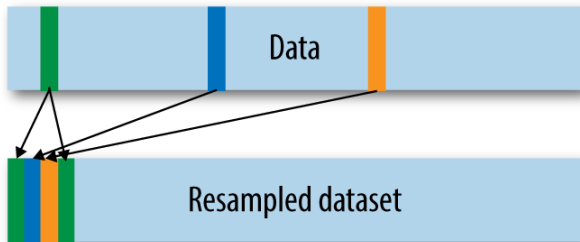


Cross-validatie

Bootstrap cross-validation

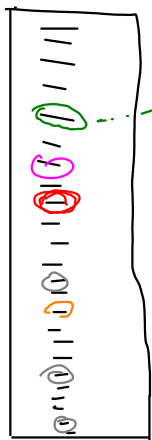
- Sampling met teruglegging
- Validatie op niet geselecteerde examples

→ zie bagging (serie 5)



Model. (c = ... gem = ...)

1000
samples



$\approx \frac{2}{3}$
TRAIN

100

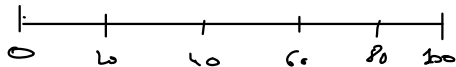


...

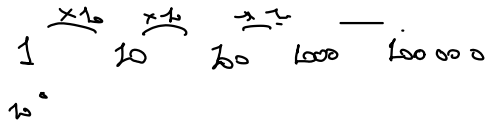
$\approx \frac{1}{3}$

↳ Evaluative

limspace (0, 100, 5)

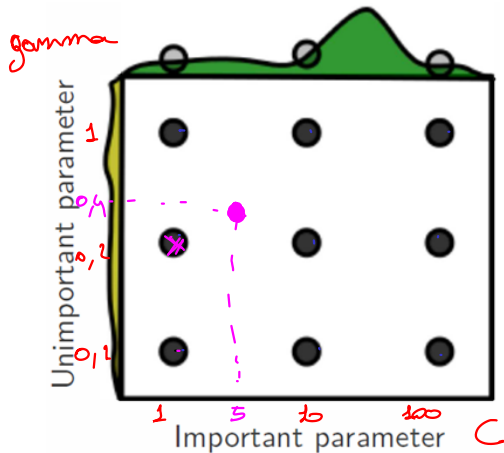


logspace (1; 100000, 5)



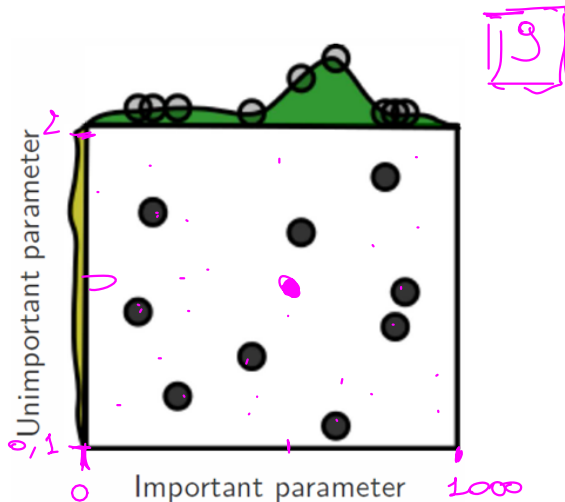
Hyperparameter tuning via cross-validation

Grid search



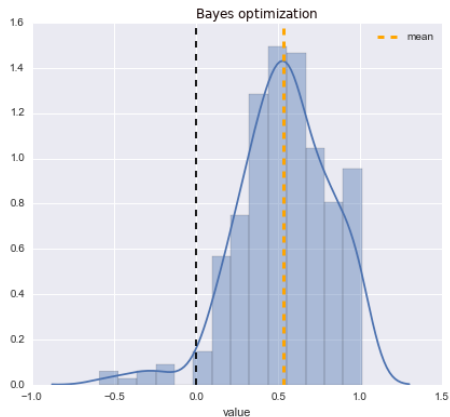
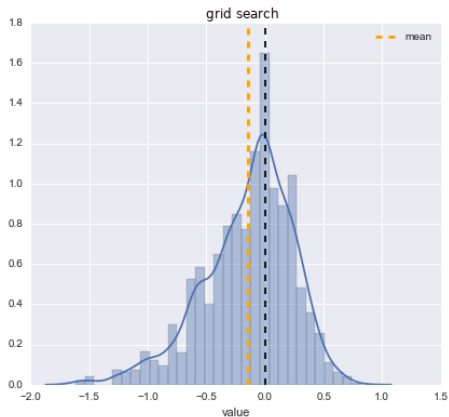
Hyperparameter tuning via cross-validation

→ Random search

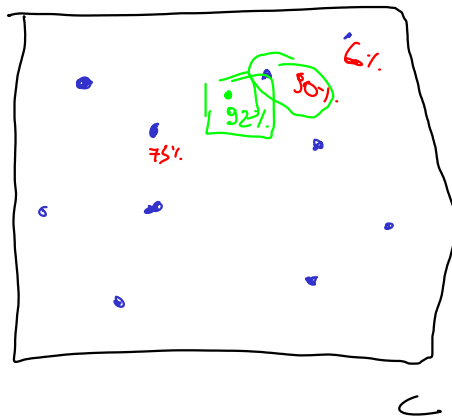


Hyperparameter tuning via cross-validation

Bayes optimization



gamm.

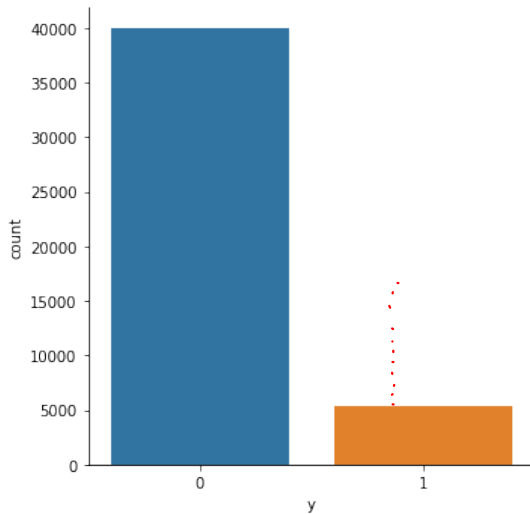
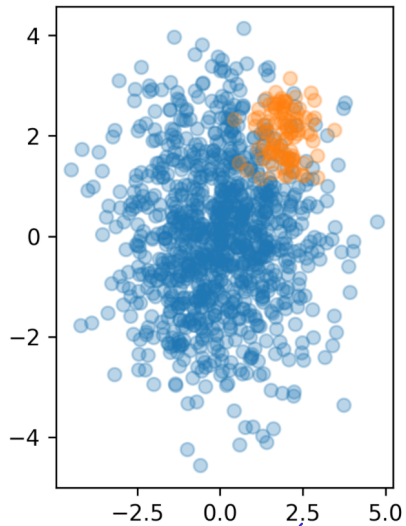


STRATIFIED SAMPLING



Niet-gebalanceerde data

Problematiek



Problematiek

	precision	recall	f1-score	support
0	0.90	0.99	0.94	8824
1	0.62	0.18	0.27	1176
micro avg	0.89	0.89	0.89	10000
macro avg	0.76	0.58	0.61	10000
weighted avg	0.87	0.89	0.86	10000

[[8694 130]
[968 208]]
89.02

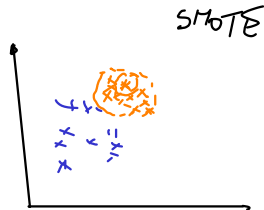
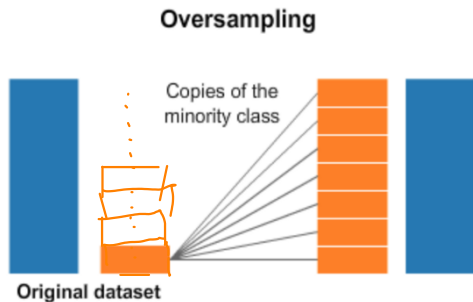
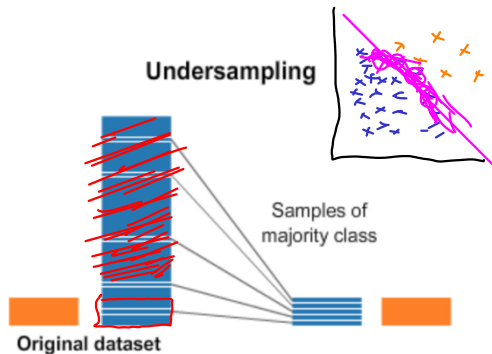
Omgaan met niet-gebalanceerde data

Verzamel meer data (van de minderheidsklasse)



Omgaan met niet-gebalanceerde data

Undersampling & Oversampling



Omgaan met niet-gebalanceerde data

Andere scoring parameter / metric kiezen

In plaats van het model te kiezen dat de hoogste accuracy oplevert, kies bijvoorbeeld het model dat de hoogste f1-score oplevert.

```
https://scikit-learn.org/stable/modules/model\_evaluation.html#the-scoring-parameter-defining-model-evaluation-rules
```

	PREDICTED		
	KAT	PAARD	HOND
TRUE KAT	4	1	1
PAARD	6	2	2
HOND	3	0	6

$$\text{Acc} = \frac{12}{25} \approx 0,48$$

$$\text{Recall}_{\text{-kat}} = \frac{4}{6} = 0,67$$

$$\text{Recall}_{\text{-paard}} = \frac{2}{20} = 0,10$$

$$\text{Recall}_{\text{-hond}} = \frac{6}{9} = 0,67$$

$$P_{1\text{-kat}} = \frac{4}{13} = 0,31$$

$$P_{1\text{-paard}} = \frac{2}{3} = 0,67$$

$$P_{1\text{-hond}} = \frac{6}{9} = 0,67$$

$$\text{Weighted recall} =$$

$$\frac{6 \times 0,67 + 1 \times 0,10 + 9 \times 0,67}{25}$$

$$\text{macro recall} = \frac{0,67 + 0,10 + 0,67}{3}$$

$$F_{1\text{-kat}} = \frac{2 \times 0,67 \times 0,31}{0,67 + 0,31} = 0,42$$

$$F_{1\text{-paard}} = \dots = 0,31$$

$$F_{1\text{-hond}} = \dots = 0,67$$

		PREDICTED		
	KAT	PAARD	MOND	
TRUE	KAT	4	1	1
	PAARD	6	2	2
	MOND	3	0	6

$$\text{min. Recall} = \frac{TP_1 + TP_2 + TP_3}{(TP_1 + TP_2 + TP_3) + (FN_1 + FN_2 + FN_3)}$$

$$= \frac{4 + 2 + 6}{4 + 2 + 6 + 2 + 8 + 3}$$

$$= \frac{12}{25} = 0,48$$

$$\text{min } f_1 = \frac{2 \mu_{Rec} \cdot \mu_{P_2}}{\mu_{Rec} + \mu_{P_3}} =$$

Acc = min Recall = min precision

Multi-class

Acc \neq min Recall \neq min precision \neq min f1

Multi-Label

β - beta score

$$f_1 = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}$$

$$\boxed{\beta = 1}$$

$$f_\beta = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}$$

$$f_{0,5} = \frac{(1 + 0,5^2) f_2 \cdot \text{Rec}}{0,5^2 \cdot \text{prec} + \text{Recall}}$$

$\beta = 0,5$
 \rightarrow men belanj on precision

$$f_2 = \frac{(1 + 4) f_1 \cdot \text{Rec}}{4 f_1 + \text{Recall}}$$

$\beta = 2 \rightarrow$ men belanj on Recall.

Omgaan met niet-gebalanceerde data

Class-weight balancing

Misclassificaties van samples uit de minderheidsklasse leiden tijdens training tot een hogere loss.

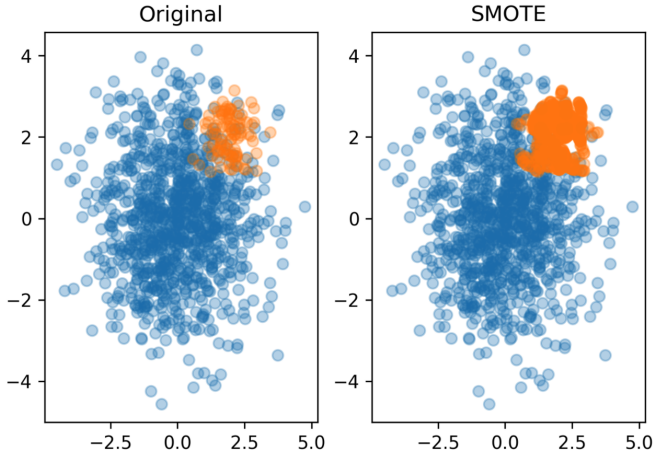
$$\text{class_weight} = \{0: 1, 1: 2\}$$

```
model = LogisticRegression(C=1, solver='liblinear', class_weight='balanced')
```

	precision	recall	f1-score	support
0	0.94	0.77	0.85	8824
→ 1	0.27	0.62	0.38	1176
micro avg	0.76	0.76	0.76	10000
macro avg	0.60	0.70	0.61	10000
weighted avg	0.86	0.76	0.79	10000
[[6830 1994]				
[444 732]]				
75.62				

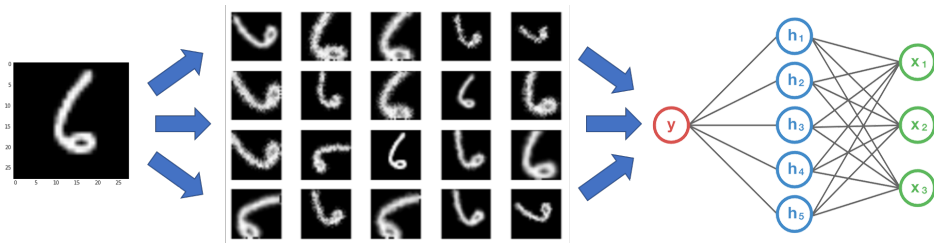
Omgaan met niet-gebalanceerde data

Data augmentation - SMOTE (Synthetic Minority Over-sampling Technique)



Omgaan met niet-gebalanceerde data

Data augmentation - image augmentation



Imbalanced Binary Classification

How to Choose A Performance Metric

