

The background features a series of concentric, light gray circles centered on the slide. Overlaid on these are stylized, light blue circuit-like lines with small circles at the ends, resembling a network or data flow, positioned in the corners.

# **SUPERVISED LEARNING - CLASSIFICATION**

JENS BAETENS

# GLOSSARY

- Supervised
- Unsupervised
- Reinforcement Learning
- Regression
- Overfitting
- Underfitting
- Learning Rate
- Loss Function
- Feature Engineering
- Normalisation
- Regularisation
- Trainen van een model

*Categoriseren → gender  
→ ouder-klasse*

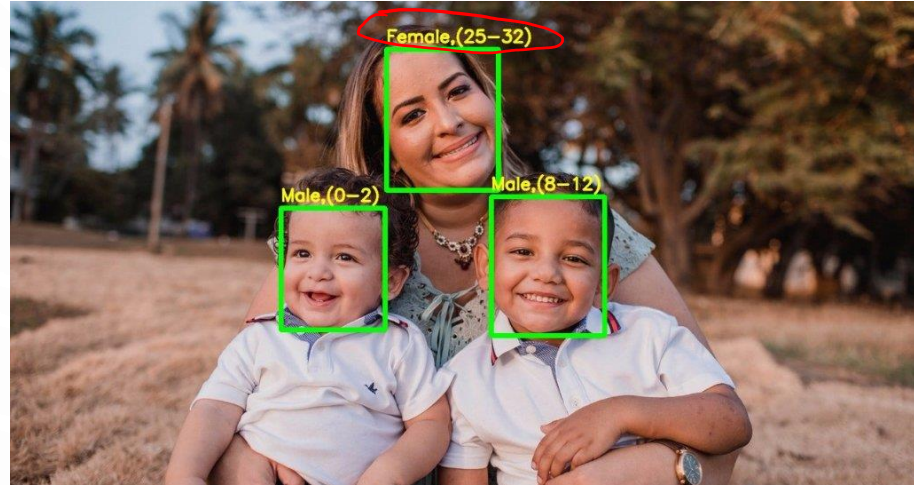
# WAT IS CLASSIFICATIE?

Supervised learning

Input omzetten naar klasse

Classifier genoemd

*AI  
model*



# WAT IS CLASSIFICATIE?

Gezichtsherkenning

Geschriftherkenning

*L-post Brussel*

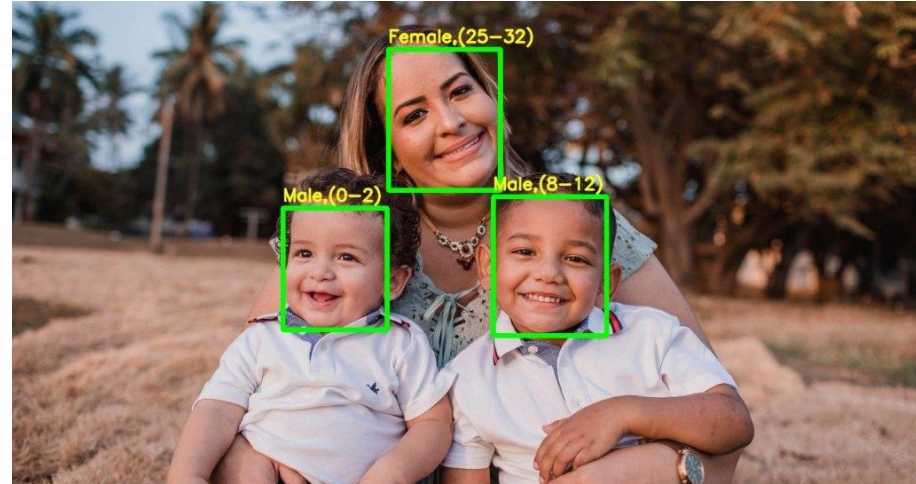
Spam detectie *BRUSSEL*

*Ja ✓ → Nee*

Kwaliteitscontroles

Medische diagnoses

*✓ → Kanker*



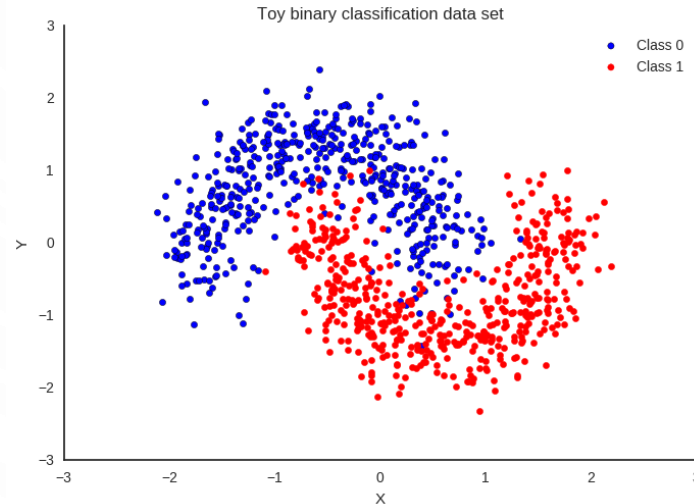
# TYPES CLASSIFIERS - BINARY

*True*  $\leftarrow$   $\rightarrow$  *False*

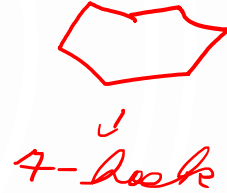
Twee verschillende klassen

Voorbeeld: Goede of slechte kwaliteit, man of vrouw, Goed- of kwaadaardig

*Basis voor de  
andere*

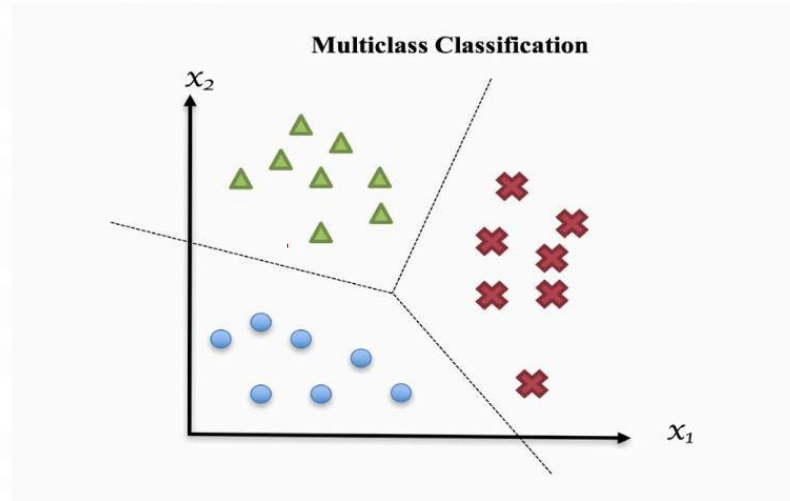


# TYPES CLASSIFIERS - MULTICLASS



$N > 2$  verschillende klassen (maar 1 mogelijk voor elke input)

Voorbeeld: Gezichtsherkenning ( 1 klasse per persoon), Hondenrasherkenning, ...



# TYPES CLASSIFIERS - MULTILABEL

B in Class 1 → Ja/Ne  
2 → Ja/Ne  
3 →  
4  
5  
:  
—  
—  
—

$N > 2$  verschillende klassen maar meerdere mogelijk per input

Voorbeeld: Beeldherkenning, Meerdere genres mogelijk voor een film, ...

Binary  
Classification



- Spam
- Not spam

Multiclass  
Classification



- Mass 1
- Dog
  - Cat
  - Horse
  - Fish
  - Bird
  - ...

Multi-label  
Classification



- Dog *Keer*
- Cat *Ja*
- Horse *Ne*
- Fish *Ne*
- Bird *Ja*
- ...

# KAN HET MET LINEAIRE REGRESSIE?

Label

$< 0,5$  Grape-Fruit  
 $> 0,5$  Orange

		name	diameter	weight	red	green	blue
grapefruit	9995	grapefruit 0	15.35	253.89	149	77	20
	9996	grapefruit 0	15.41	254.67	148	68	7
	9997	grapefruit 0	15.59	256.50	168	82	20
	9998	grapefruit 0	15.92	260.14	142	72	11
	9999	grapefruit 0	16.45	<u>261.51</u>	152	74	2
orange	0	orange 1	2.96	<u>86.76</u>	172	85	2
	1	orange 1	3.91	88.05	166	78	3
	2	orange 1	4.42	95.17	156	81	2
	3	orange 1	4.47	95.60	163	81	4
	4	orange 1	4.48	95.76	161	72	9



→ *enkele gebied vrij groot*

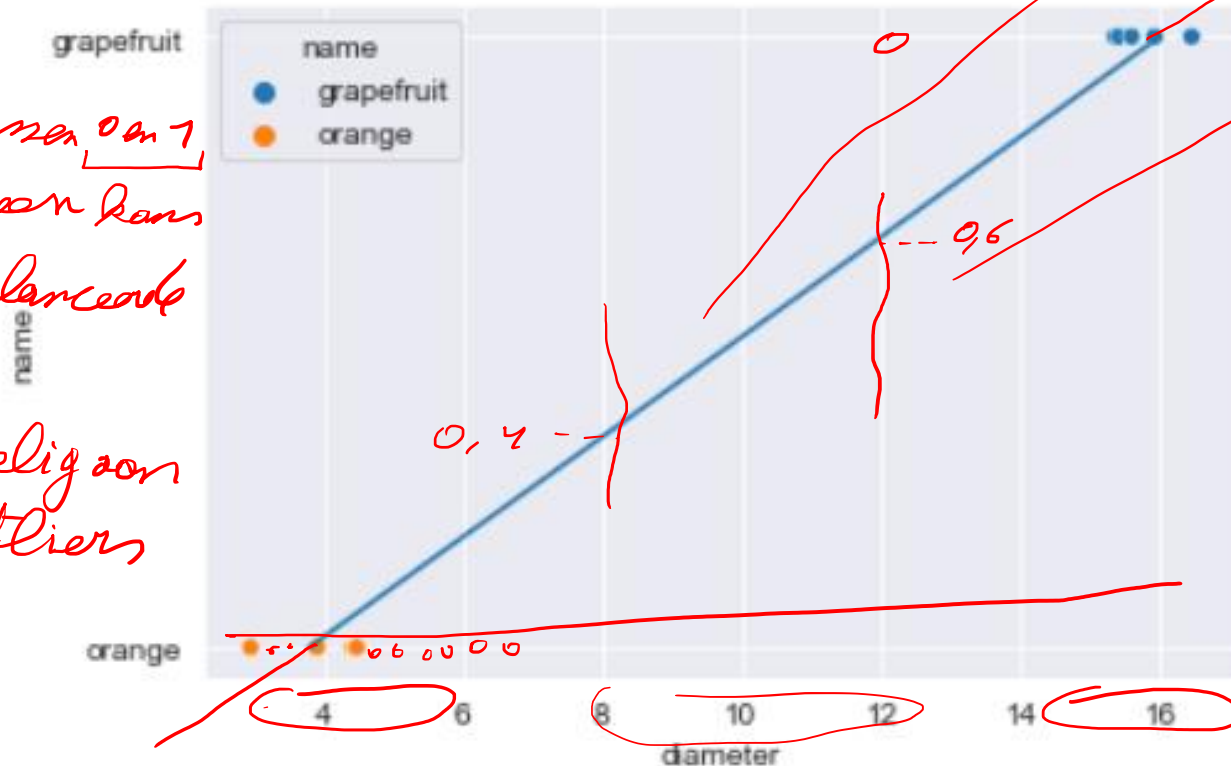
## KAN HET MET LINEAIRE REGRESSIE?

→ *niet kunnen aan 1*

→ *geen kans*

→ *ongebalanceerde data*

→ *gevoelig aan outliers*

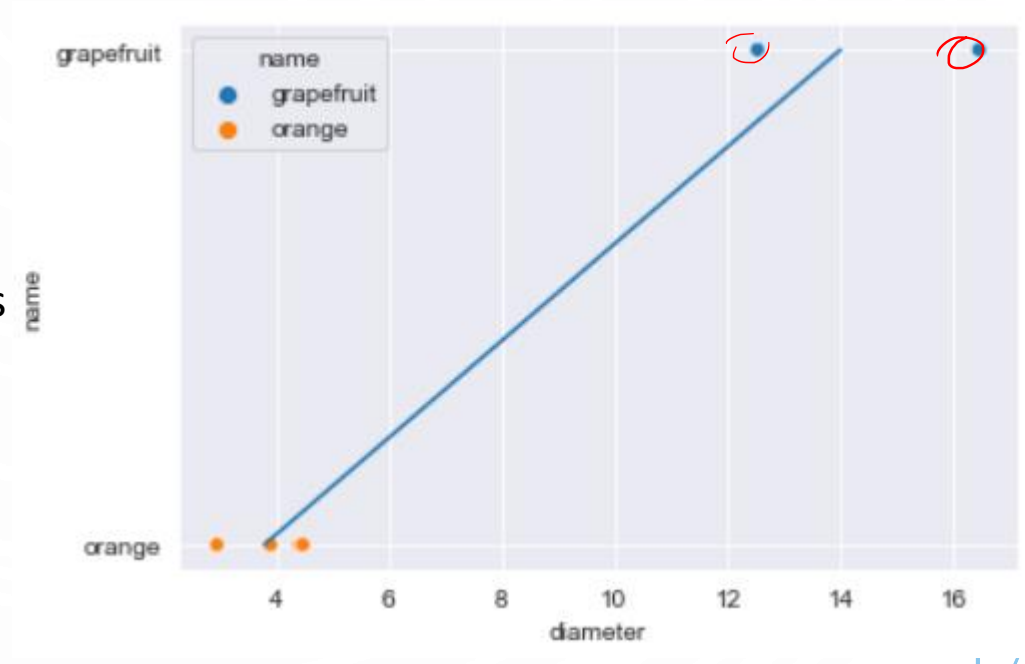


# KAN HET MET LINEAIRE REGRESSIE?

Gevoelig voor outliers

Zeer breed “fuzzy” middenstuk

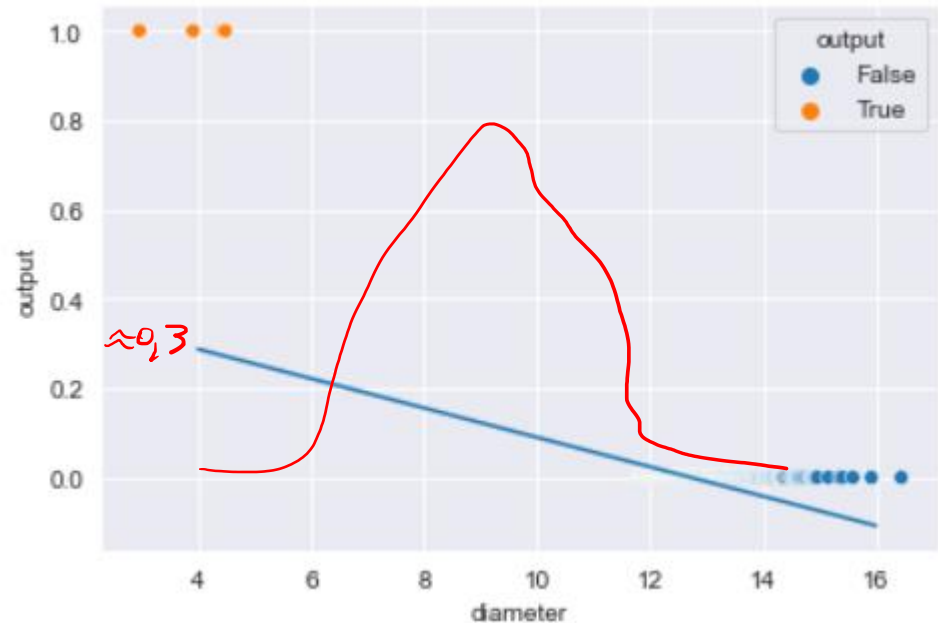
Komt niet overeen met een kans



# KAN HET MET LINEAIRE REGRESSIE?

Ongebalanceerde klassen

=> Geen lineaire regressie mogelijk



# CLASSIFICATIE – LOGISTIC REGRESSION

JENS BAETENS

# LOGISTIC REGRESSION

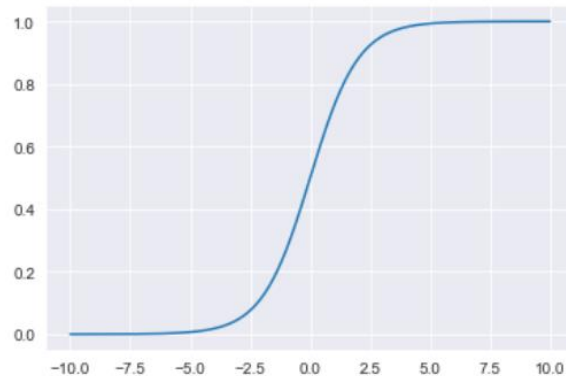
Logistische functie (sigmoid)

$$f(z) = \frac{1}{1+e^{-z}}$$

Geeft een waarde terug tussen 0 en 1  
- De kans het tot de klasse hoort

$$f_w(x) = \frac{1}{1+e^{-w^T x}}$$

$$w^T x = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_N x_N \approx \text{lineaire regressie}$$



# LOGISTIC REGRESSION

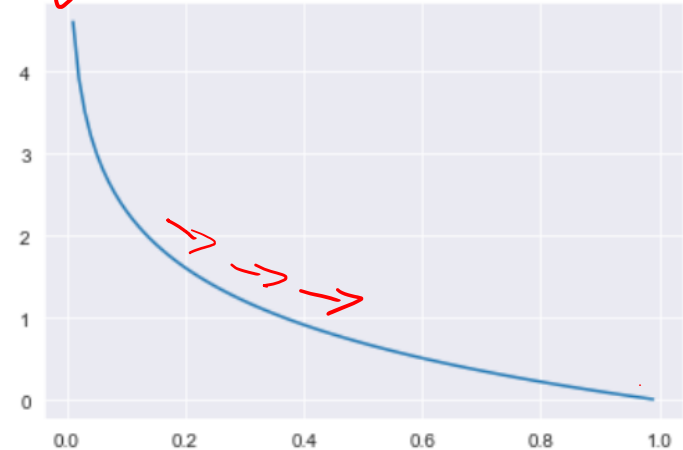
$$L(\mathbf{w}) = \begin{cases} -\ln(f_{\mathbf{w}}(x)) & \text{als } y = 1 \\ -\ln(1 - f_{\mathbf{w}}(x)) & \text{als } y = 0 \end{cases}$$

$$L(\mathbf{w}) = -\frac{1}{N} \left[ \sum_{i=1}^N y_i \ln(f_{\mathbf{w}}(x_i)) + (1 - y_i) \ln(1 - f_{\mathbf{w}}(x_i)) \right]$$

*→ alle bijl maar 1 term*

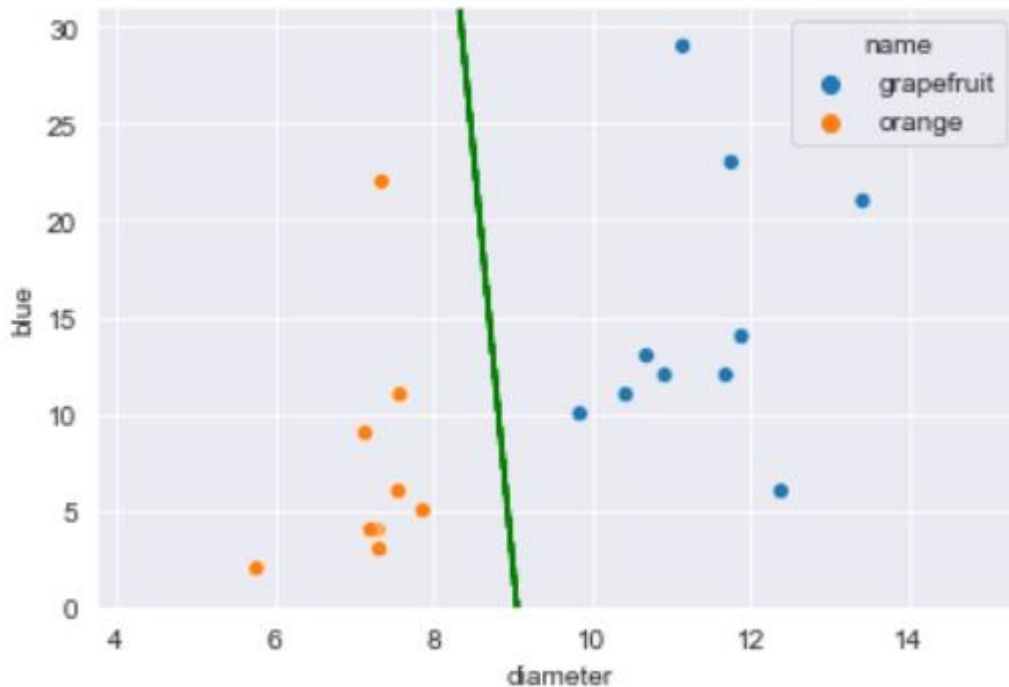
Minimalisatie dmv Gradient Descent

*mis = hoge fout*



# LOGISTIC REGRESSION

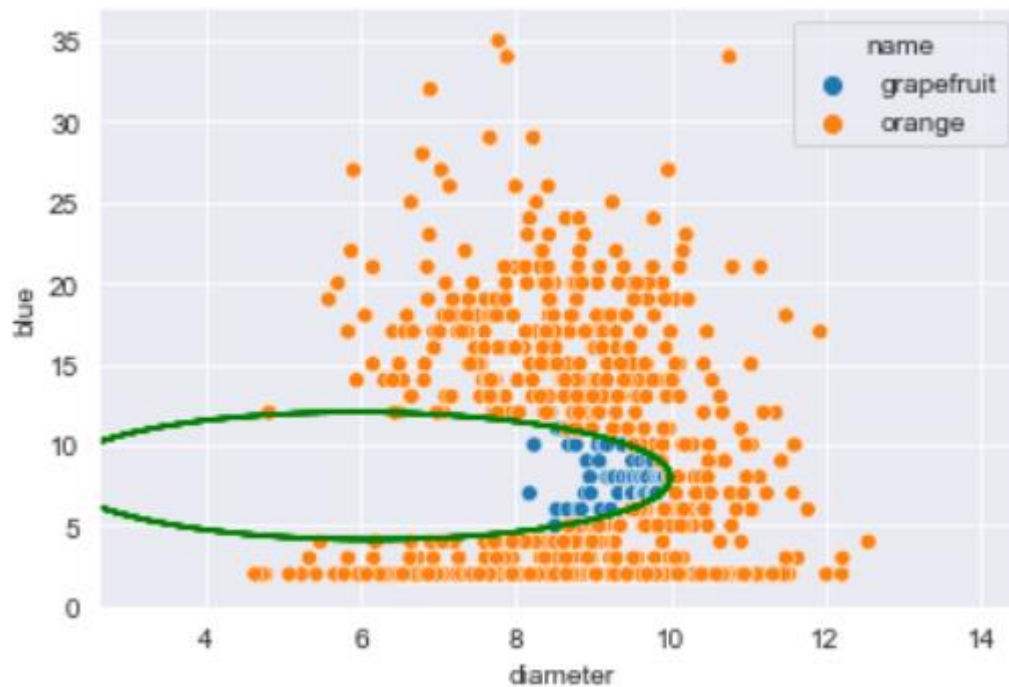
Links v.d. lijn = orange  
↗  
rechts v.d. lijn = sinaasappel.



[illegible]



A scatter plot showing the relationship between 'diameter' (x-axis) and 'blue' (y-axis) for two fruit types: 'grapefruit' (blue dots) and 'orange' (orange dots). The x-axis ranges from 4 to 14, and the y-axis ranges from 0 to 35. The 'orange' data points are widely scattered, mostly between diameter 6 and 12 and blue 0 and 35. The 'grapefruit' data points are clustered together, mostly between diameter 8 and 10 and blue 5 and 10. A green ellipse is drawn around the 'grapefruit' cluster, indicating a region of interest.

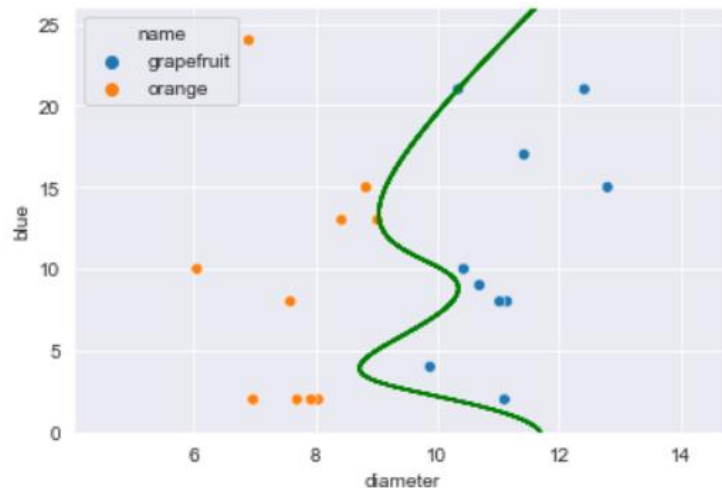


# LOGISTIC REGRESSION – REGULARISATIE

Regularisatie via C-parameter

Inverse regularisatie sterkte

Hoge waarde = weinig regularisatie



```
model = LogisticRegression(C=10) # C= inverse regularisatiesterkte  
model.fit(X, df_trimmed.output)
```

# LOGISTIC REGRESSION – EVALUATIE

Accuraatheid  $\frac{98}{100} = 98\%$

Precisie  $\frac{8}{908}$   $\frac{TP}{TP+FP}$

Recall-Specificiteit  $\frac{TN}{TN+FP}$

Recall  $80\%$   $\frac{TP}{TP+FN}$

F1-Score  $2 \frac{Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP+FP+FN}$

Recall+ Recall-  
↓ ↓  
Vijver/Daten

True Class

10 Positive 990 Negative

		True Class	
		Positive	Negative
Predicted Class	Positive	TP 8	FP 900
	Negative	FN 908	TN 90

Precisie  
→  
Test  
Precisie  
→

# VOORBEELD CONFUSION MATRIX

8 vissen en 4 flessen

Recall  
Sensitiviteit  
↓ 6/8

Recall  
Specificiteit  
↓ 3/4

Accuraatheid = 9/12

6 ~~vis~~ + 3 Fl

Vissen ✓

Sensitiviteit/Recall = 6/8

6 Vissen

2 flessen

Totaal Vissen

- Weinig positieve samples gemist

Specificiteit = 3/4 (enkel voor 2 flessen)

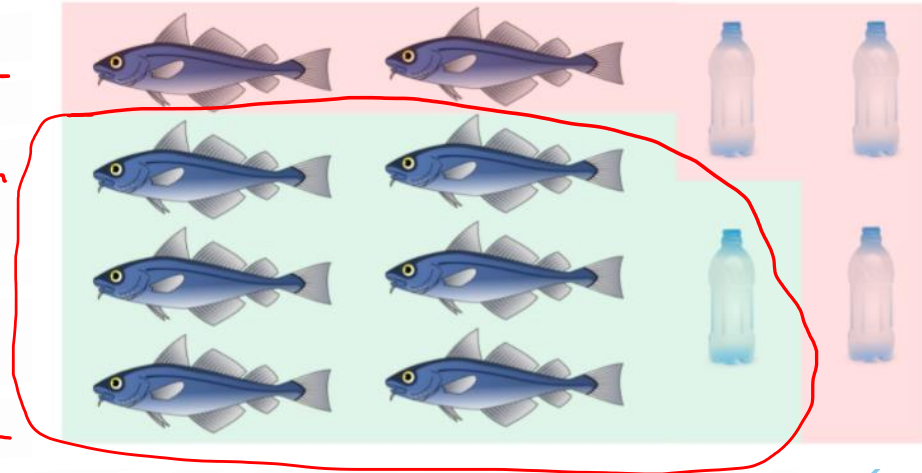
- Weinig negatieve samples gemist

Precision = 6/7

Voorkaas

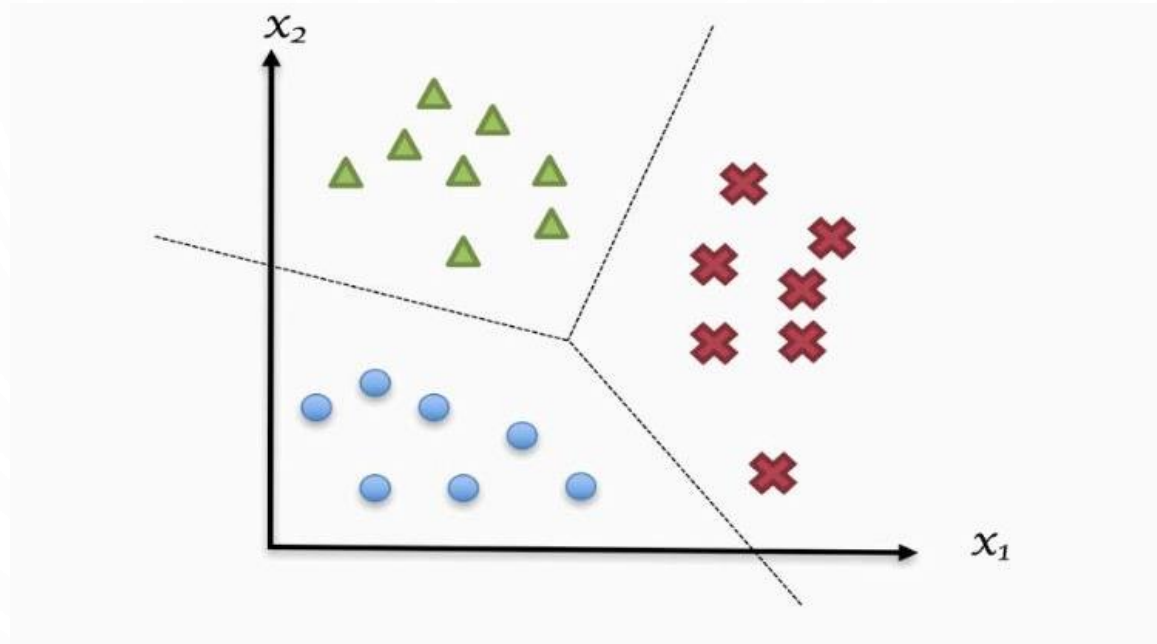
- Weinig negatieve samples als positieve geclassificeerd

Fl als vis classificeren



Voorbeeld (ML → vis classificeren)

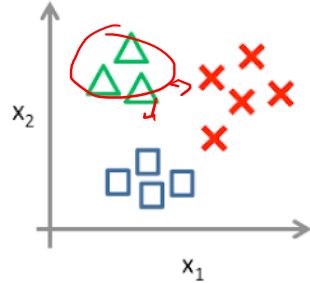
# LOGISTIC REGRESSION – MULTICLASS



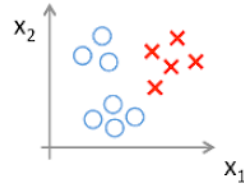
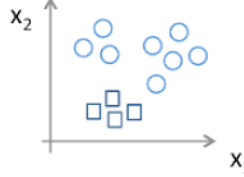
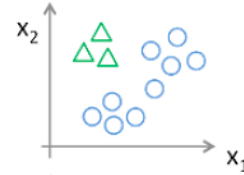
Multi-class classificatie bestaat uit meerdere

## LOGISTIC REGRESSION – ONE VS ALL *binair classificatie*

One-vs-all (one-vs-rest):



Class 1: **Green**  
Class 2: **Blue**  
Class 3: **Red**

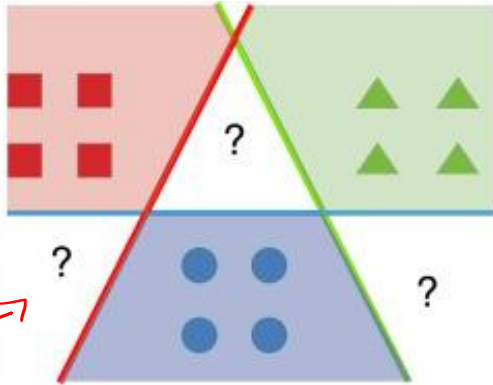


*N classificatie nodig  
1/klass*

*Conflicten mogelijk  
→ Kies met de  
hoogste kans*

# LOGISTIC REGRESSION – ONE VS ONE

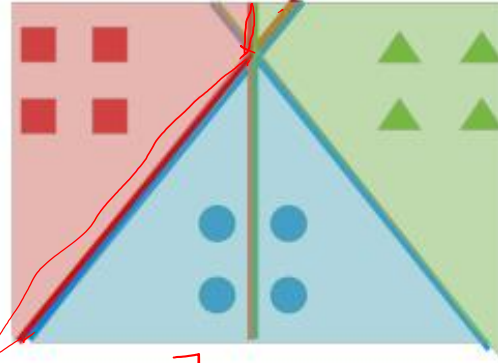
One vs all



Conflicten →

Is het rood?

→ One vs One



Is het niet rood?

- R-B
- R-G
- G-B

$N(N-1)$  classificaties nodig  
→ trainingstijd is langer.

# LOGISTIC REGRESSION – MULTICLASS EVALUATION

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1



# LOGISTIC REGRESSION – MULTICLASS EVALUATION

*Precision* →

*Recall* ↓

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

*Micro → total elements of Macro*

*Weighted*

Class	Precision	Recall	F1-score
Apple	$7/24$ 0.29	0.64	0.40
Orange	0.33	0.17	0.22
Mango	0.17	0.08 $1/12$	0.11

*Micro*  $10/36$   $10/36$   $10/36$

*Macro*  $\frac{0.29 + 0.33 + 0.17}{3}$   $\frac{0.64 + 0.17 + 0.08}{3}$  ...

*Weighted*  $\frac{0.29 \cdot 24 + 0.33 \cdot 12 + 0.17 \cdot 6}{36}$

# LOGISTIC REGRESSION – MULTICLASS EVALUATIE

Micro – F1: Globale waarden

Macro – F1: Gemiddelde F1 – scores

Weighted F1: Gew. Gemiddelde  
- Gewichten = # samples

*in de data  
niet in de voorbeelden*

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

Class	Precision	Recall	F1-score
Apple	0.29	0.64	0.40
Orange	0.33	0.17	0.22
Mango	0.17	0.08	0.11

# GLOSSARY

- Classificatie
- Binary classifier
- Multi-class classifier
- Multi-label classifier
- True/False Positive/Negative
- Accuraatheid / Specificiteit / ...
- One-vs-All
- One-vs-One
- Confusion matrix