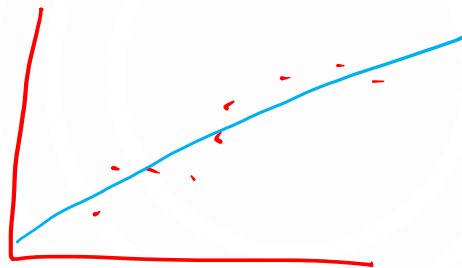
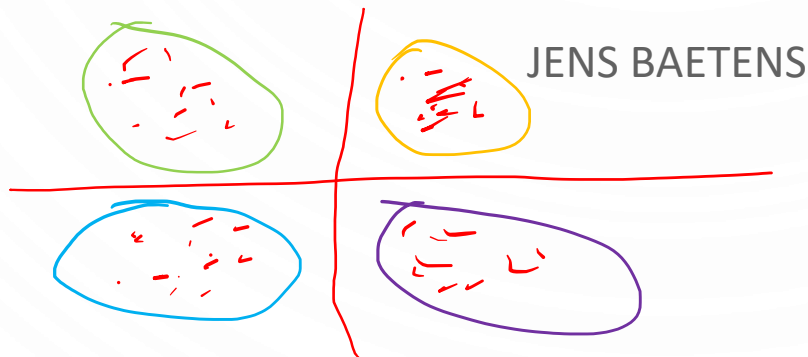


regressie




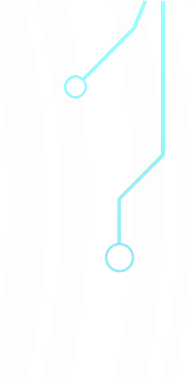
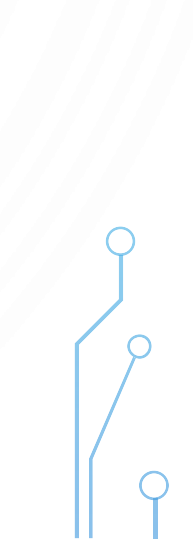
SUPERVISED LEARNING - CLASSIFICATION

4 classes → categoriser





GLOSSARY

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

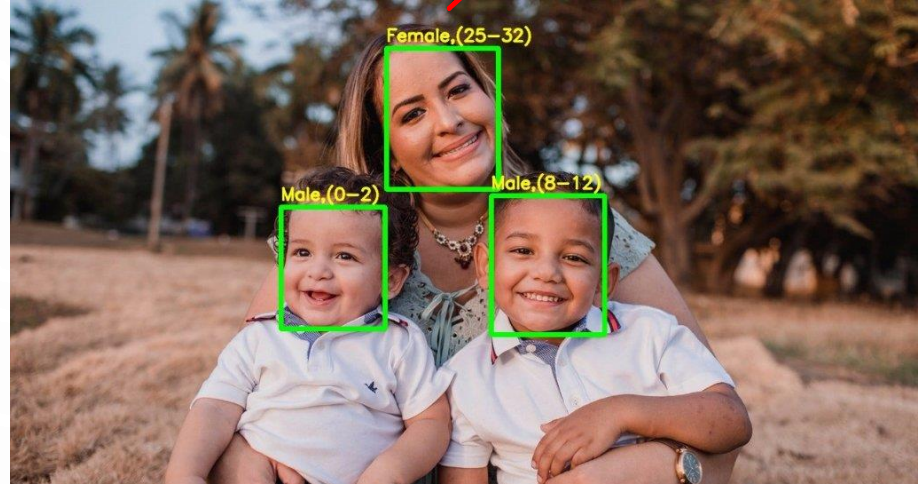
WAT IS CLASSIFICATIE?

features/inputs → *labels/targets/outputs*

Supervised learning

Input omzetten naar klasse

Classifier genoemd



+

Regressor

= model dat
regressie doet

WAT IS CLASSIFICATIE?

Gezichtsherkenning

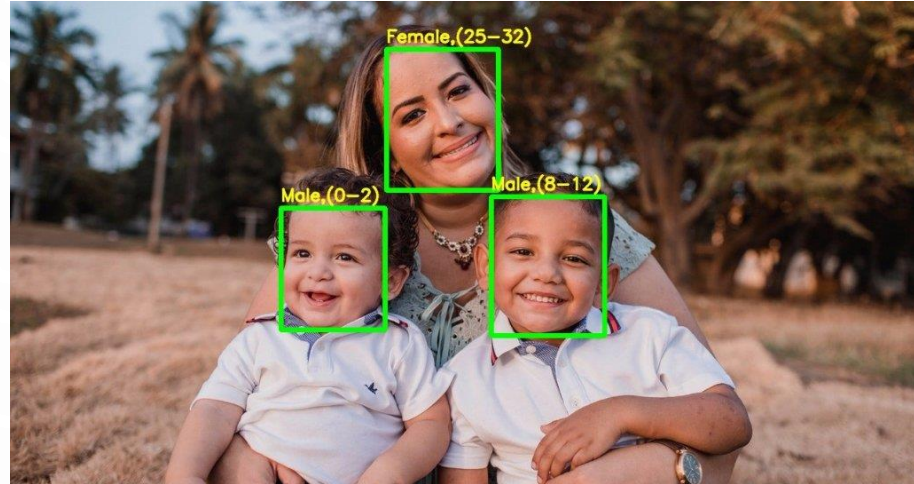
Geschriftherkenning

6 → 6 7 = 7

Spam detectie

Kwaliteitscontroles

Medische diagnoses



TYPES CLASSIFIERS - BINARY

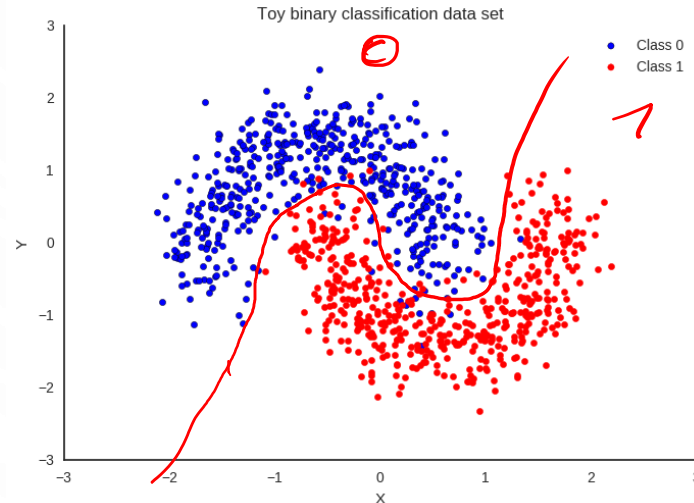
Twee verschillende klassen

klase A of niet

klase A of klase B

Ja

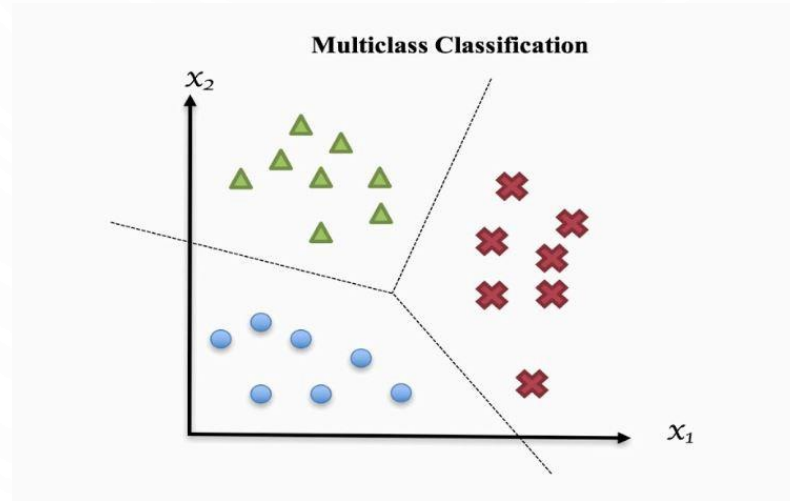
Voorbeeld: Goede of slechte kwaliteit, man of vrouw, Goed- of kwaadaardig



TYPES CLASSIFIERS - MULTICLASS

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

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



TYPES CLASSIFIERS - MULTILABEL

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

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

Voorbeeld van n.v.t.

Binary
Classification



- Spam
- Not spam

Multiclass
Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

Multi-label
Classification

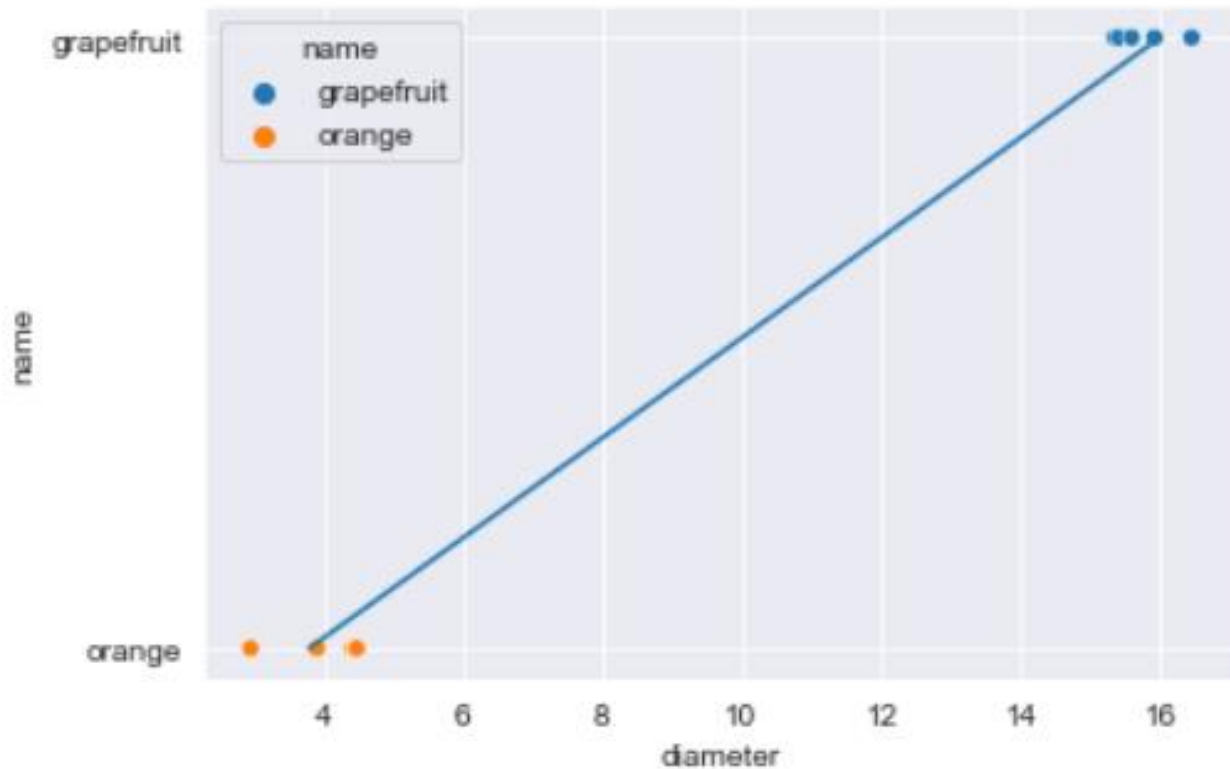


- Dog
- Cat
- Horse
- Fish
- Bird
- ...

KAN HET MET LINEAIRE REGRESSIE?

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

KAN HET MET LINEAIRE REGRESSIE?

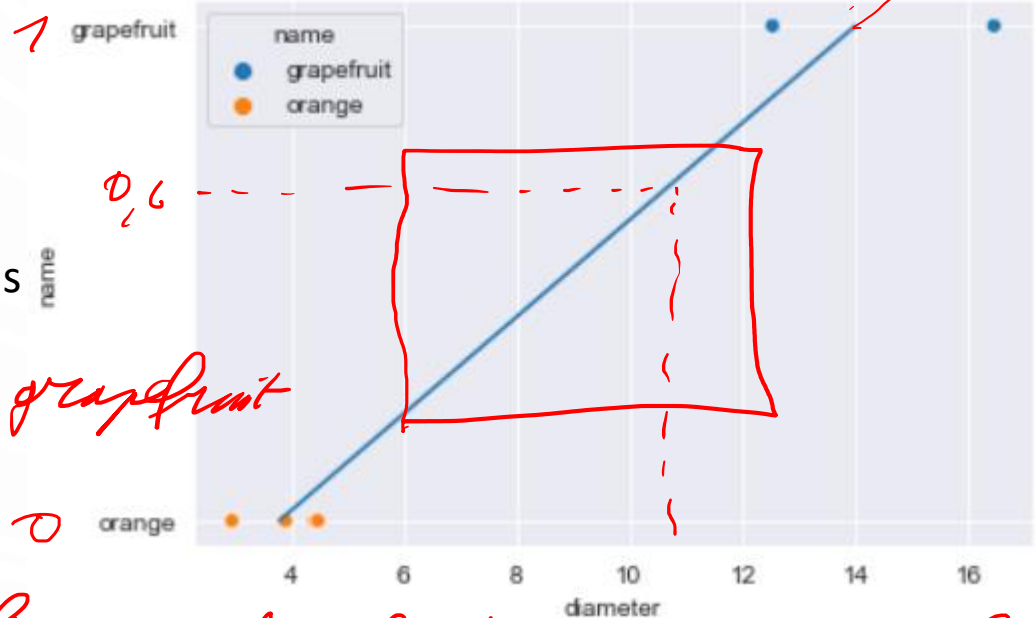


KAN HET MET LINEAIRE REGRESSIE?

Gevoelig voor outliers

Zeer breed “fuzzy” middenstuk

Komt niet overeen met een kans



0,6 → kans dat grapefruit is

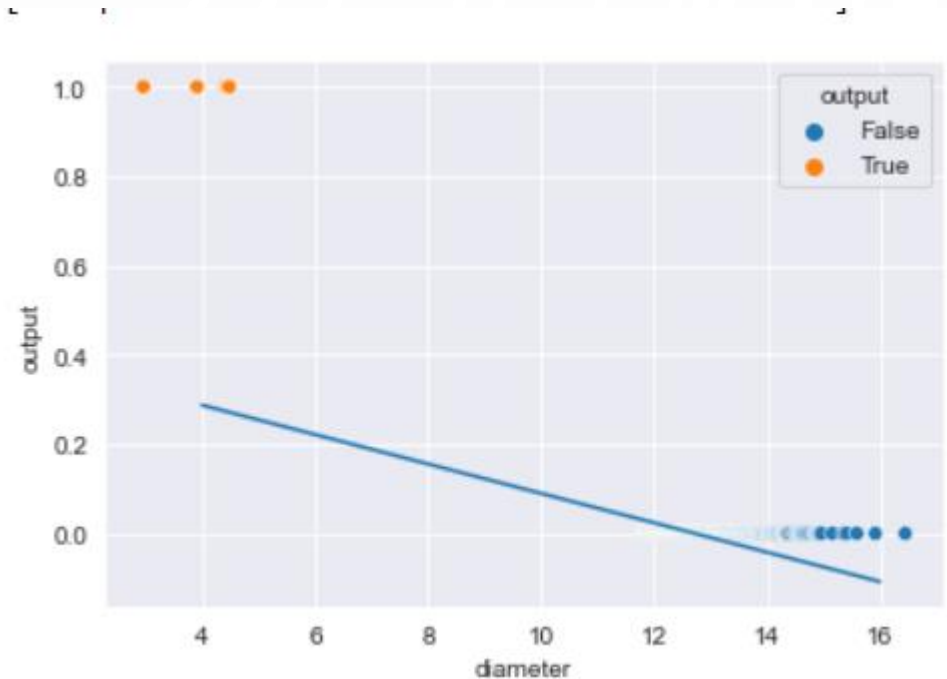
↳ tussen 0 en 1

→ belangrijk om zekerheid van het model te weten

KAN HET MET LINEAIRE REGRESSIE?

Ongebalanceerde klassen

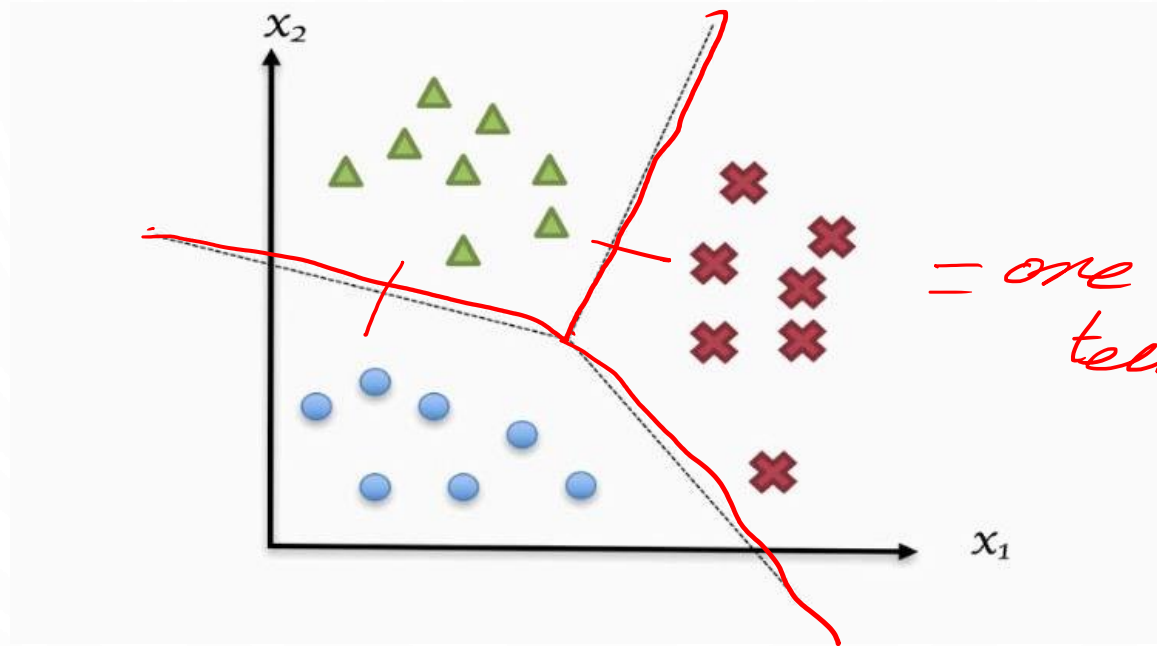
=> Geen lineaire regressie mogelijk



CLASSIFICATION – MULTICLASS

one vs one

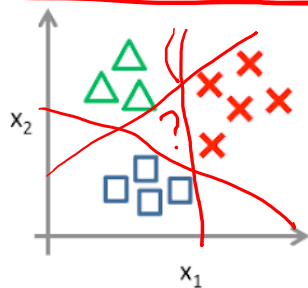
one vs all



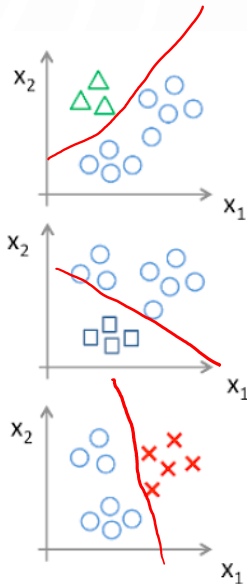
= one vs one
technique

CLASSIFICATION – ONE VS ALL

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



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



N classes
↓
N classifiers

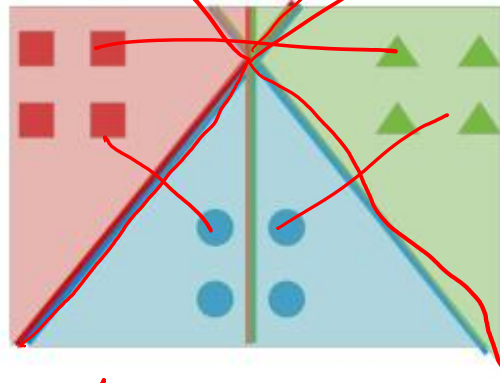
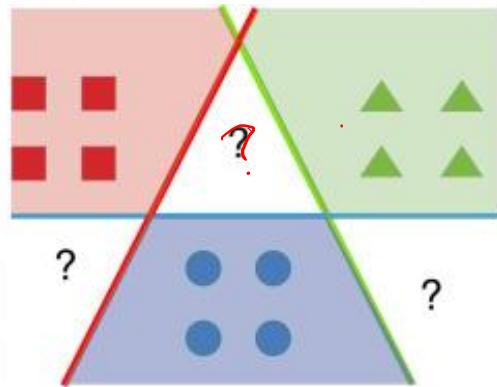
→ ~~one-vs-one~~
→ gemakkelijker te trainen
→ niet altijd een antwoord

("multilabel")

multiclass classifier

CLASSIFICATION – ONE VS ONE

N classes $\rightarrow \frac{N \cdot (N-1)}{2}$ classifiers nodig



\rightarrow altijd een antwoord
 \rightarrow maar men classifier nodig

Evaluatie

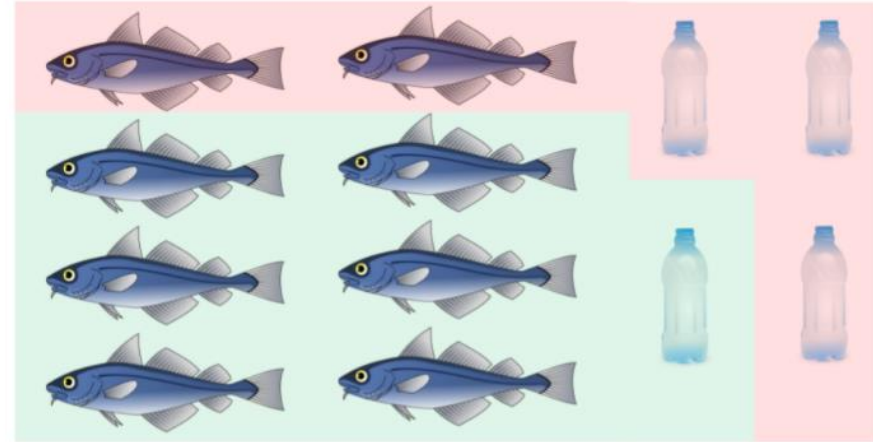
VOORBEELD CONFUSION MATRIX

Je wilt vissen uit een vijver waar ook flessen in drijven

- In totaal 8 vissen en 4 flessen

Na het vissen heb je ~~8~~₆ vissen en 1 fles opgevist

Hoe goed is je resultaat?



CLASSIFICATIE – EVALUATIE

Accuraatheid

$$\frac{9}{12}$$

Precisie

$$\frac{TP}{TP+FP}$$

Specificiteit

$$\frac{TN}{TN+FP}$$

Recall

$$\frac{TP}{TP+FN}$$

F1-Score

$$2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP+FP+FN}$$

confusion matrix

11

8 vinnen

4 flessen

True Class

Positive

Negative

Predicted Class
Positive
Negative

	Positive	Negative
Positive	6 TP	1 FP
Negative	2 FN	3 TN

VOORBEELD CONFUSION MATRIX

Accuraatheid = $9/12$ *6 vissen + 3 flessen*

Sensitiviteit/Recall = $6/8$ *→ Portiere*

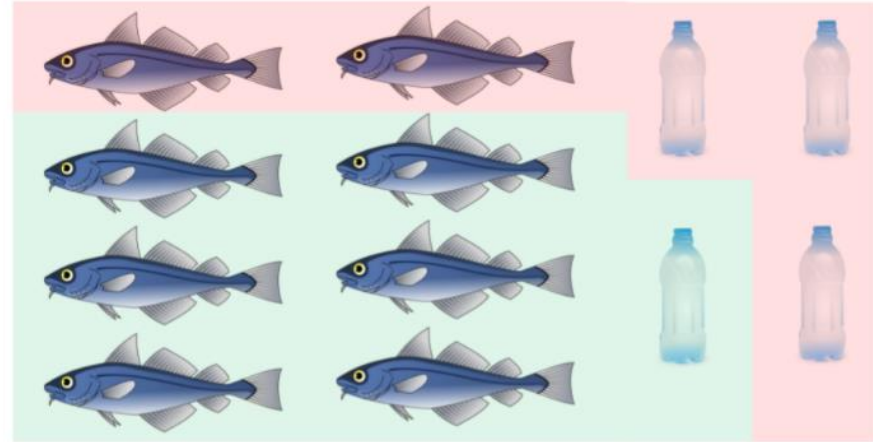
- Weinig positieve samples gemist

Specificiteit = $3/4$ *→ negatieve*

- Weinig negatieve samples gemist

Precision = $6/7$ *→ in de positief voorspelde raten er zoveel goede*

- Weinig negatieve samples als positieve geclassificeerd



WANNEER WELKE METRIEK GEBRUIKEN?

- Precision
 - Als de kost van een false positive hoog is
 - Spam detectie: verlies van een mail is erger dan een mogelijke spam mail doorlaten
- Recall
 - Als de kost van een false negative hoog is
 - Fraude detecteren of corona testen: Het is beter er een aantal te veel te detecteren dan besmettingen te negeren

WANNEER WELKE METRIEK GEBRUIKEN?

- Accuracy

- Kijkt naar alle voorbeelden
- Combinatie van precision en recall maar houdt geen rekening met gebalanceerdheid van de dataset

Kanker test → 9,1% heeft kanker

- F1-score

- Combinatie van precision en recall rekening houdend met de gebalanceerdheid van de dataset
- Gebruik deze als je dataset niet gebalanceerd is en zowel precision als recall belangrijk is

→ test zegt altijd geen kanker → acc 99,9%

CLASSIFICATION – MULTICLASS EVALUATION

Dataset

→ 11 apples

→ 12 oranges

→ 13 mangoes

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

$$acc = \frac{10}{36}$$

CLASSIFICATION – MULTICLASS EVALUATION

Precision →

Recall ↓

True Class

	Apple	Orange	Mango
Apple	7 <i>TP</i>	8 <i>FP</i>	9 <i>FN</i>
Orange	1	2	3
Mango	3 <i>FN</i>	2	1

Predicted Class

Class	Precision	Recall	F1-score
Apple	$\frac{7}{12} 0.29$	$\frac{7}{17} 0.64$	0.40
Orange	$\frac{2}{6} 0.33$	$\frac{2}{12} 0.17$	0.22
Mango	$\frac{1}{6} 0.17$	$\frac{1}{13} 0.08$	0.11

Per class

↳ global: *micro*
macro
weighted

CLASSIFICATION – MULTICLASS EVALUATIE

Micro-F1 (Precision/Recall)

10	26
26	20

Bereken totaal aantal TP/TN/...

Bereken hieruit de nodige metrieken

- $Total\ TP = (7+2+1) = 10$
- $Total\ FP = (8+9)+(1+3)+(3+2) = 26$
- $Total\ FN = (1+3)+(8+2)+(9+3) = 26$

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

Doordat er geen verschil is tussen P en N bij Micro scores

Precision = Recall = Micro F1 = Accuracy

CLASSIFICATION – MULTICLASS EVALUATIE

0,257
11

Macro-F1 (Precision/Recall)

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

Bereken per klasse TP/TN/...

Neem gemiddelde van de scores van elke klasse

Gemiddelde is unweighted

Voor Apple:

- $TP = 7$
- $TN = (2+3+2+1) = 8$
- $FP = (8+9) = 17$
- $FN = (1+3) = 4$

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

CLASSIFICATION – MULTICLASS EVALUATION

Weighted-F1 (Precision/Recall)

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

Weighted average in macro F1

$$\frac{11 \cdot 0,4 + 12 \cdot 0,22 + 13 \cdot 0,11}{36}$$

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

CLASSIFICATION – MULTICLASS EVALUATIE

Micro – F1: Globale waarden

Macro – F1: Gemiddelde F1 – scores

Weighted F1: Gew. Gemiddelde
- Gewichten = # samples

		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