
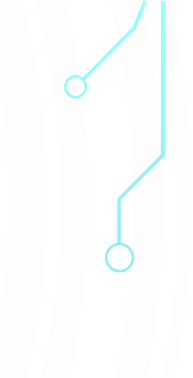
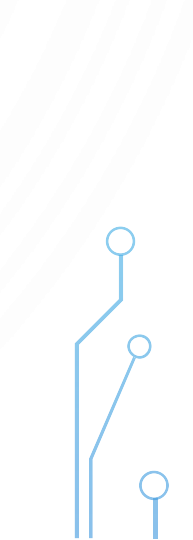


SUPERVISED LEARNING - CLASSIFICATION

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

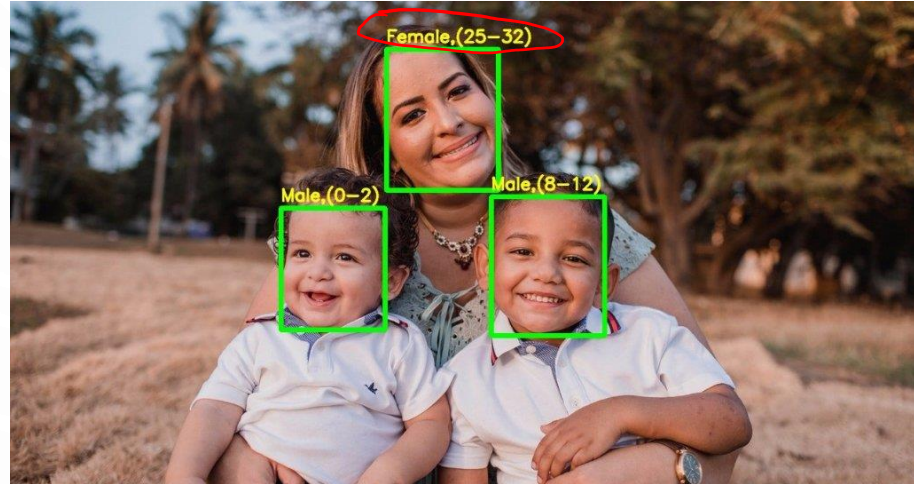
WAT IS CLASSIFICATIE?

Supervised learning

Input omzetten naar klasse

Classifier genoemd

11
model



WAT IS CLASSIFICATIE?

Gezichtsherkenning

Geschriftherkenning

L-post Brussel

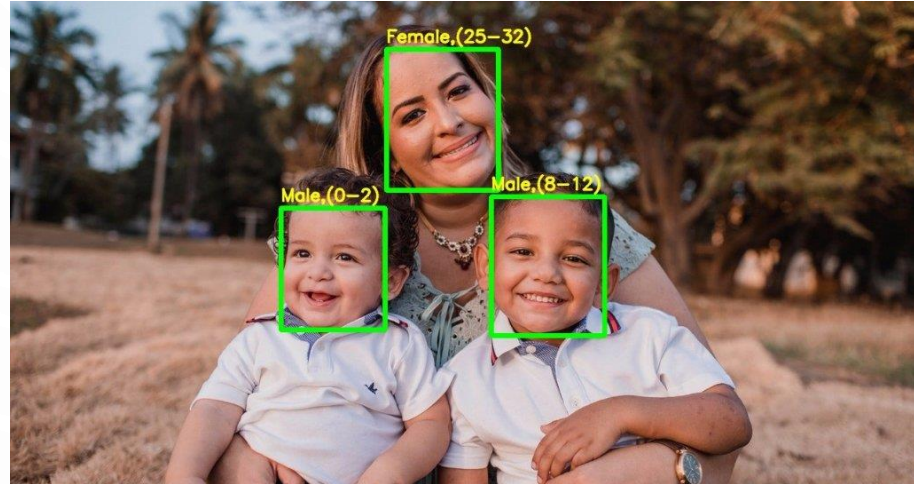
Spam detectie *BRUSSEL*

Ja ✓ → Nee

Kwaliteitscontroles

Medische diagnoses

✓ → Kanker



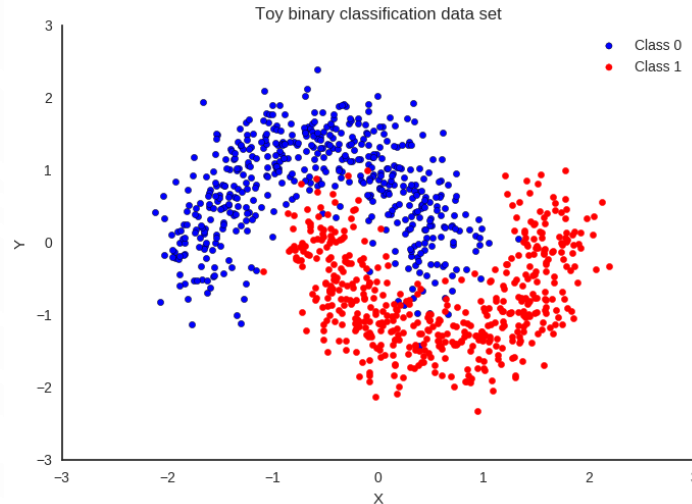
TYPES CLASSIFIERS - BINARY

True \hookrightarrow *False*

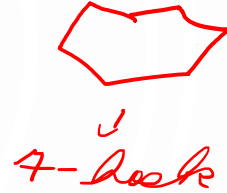
Twee verschillende klassen

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

*Basis voor de
andere*

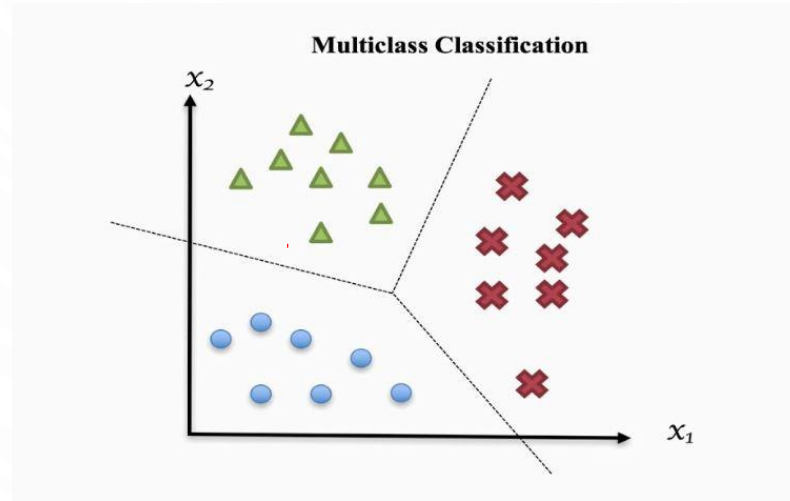


TYPES CLASSIFIERS - MULTICLASS



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

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



TYPES CLASSIFIERS - MULTILABEL

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

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

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

Binary
Classification



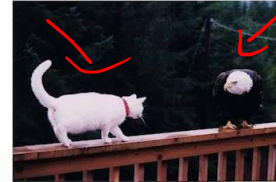
- Spam
- Not spam

Multiclass
Classification



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

Multi-label
Classification



- Dog *Kee*
- Cat *Ja*
- Horse *Nee*
- Fish *Nee*
- 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*

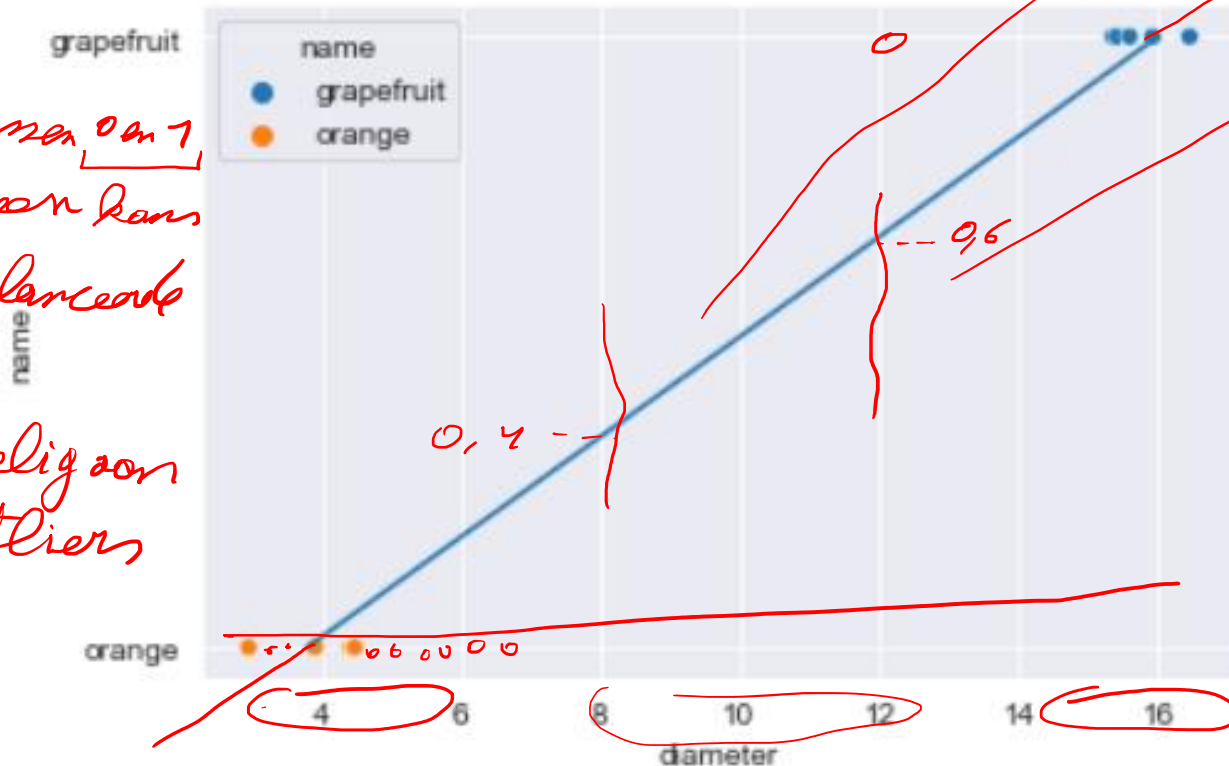
KAN HET MET LINEAIRE REGRESSIE?

→ *niet kunnen aan 1*

→ *geen kans*

→ *ongebalanceerde data*

→ *gevoelig aan outliers*

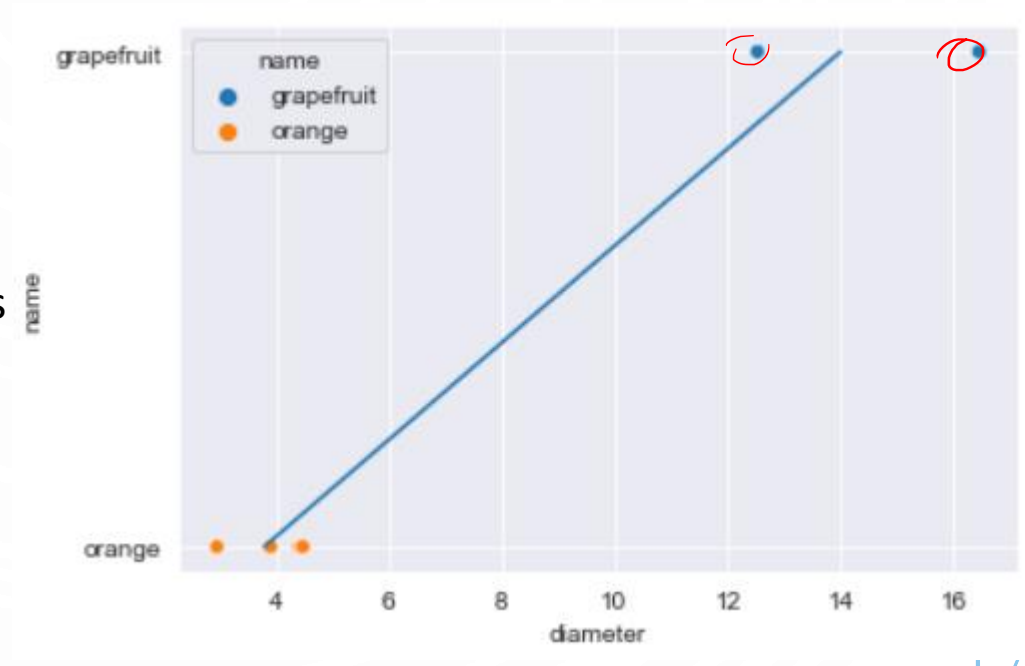


KAN HET MET LINEAIRE REGRESSIE?

Gevoelig voor outliers

Zeer breed “fuzzy” middenstuk

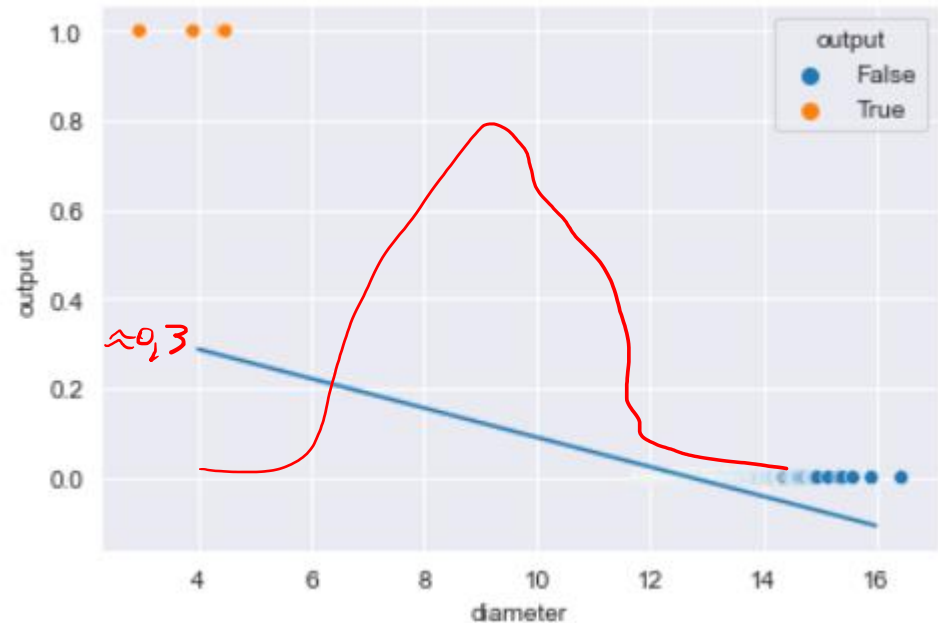
Komt niet overeen met een kans



KAN HET MET LINEAIRE REGRESSIE?

Ongebalanceerde klassen

=> Geen lineaire regressie mogelijk



CLASSIFICATIE – LOGISTIC REGRESSION

JENS BAETENS

LOGISTIC REGRESSION

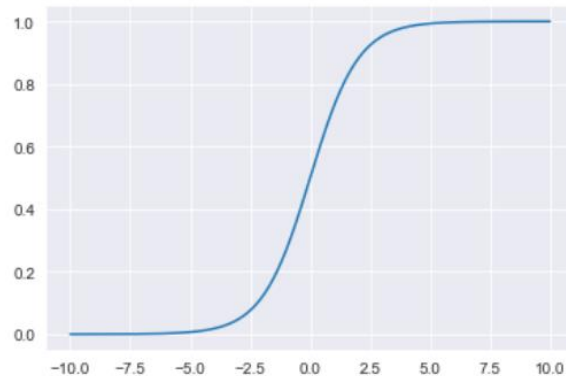
Logistische functie (sigmoid)

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

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

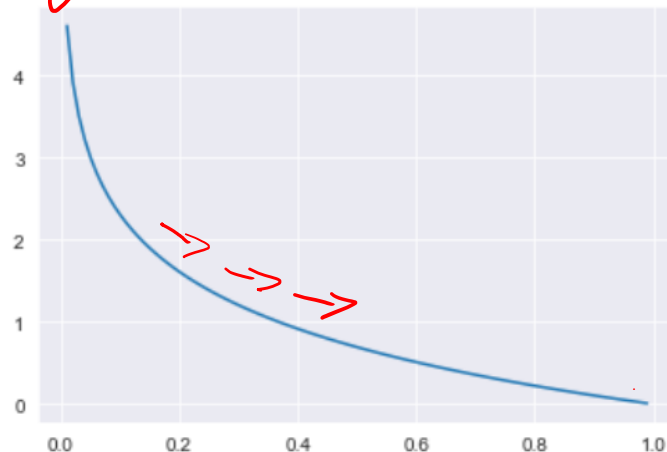
$$L(\mathbf{w}) = \begin{cases} -\ln(f_{\mathbf{w}}(x)) & \text{als } y = 1 \\ -\ln(1 - f_{\mathbf{w}}(x)) & \text{als } y = 0 \end{cases}$$

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

→ alle bijl maar 1 term

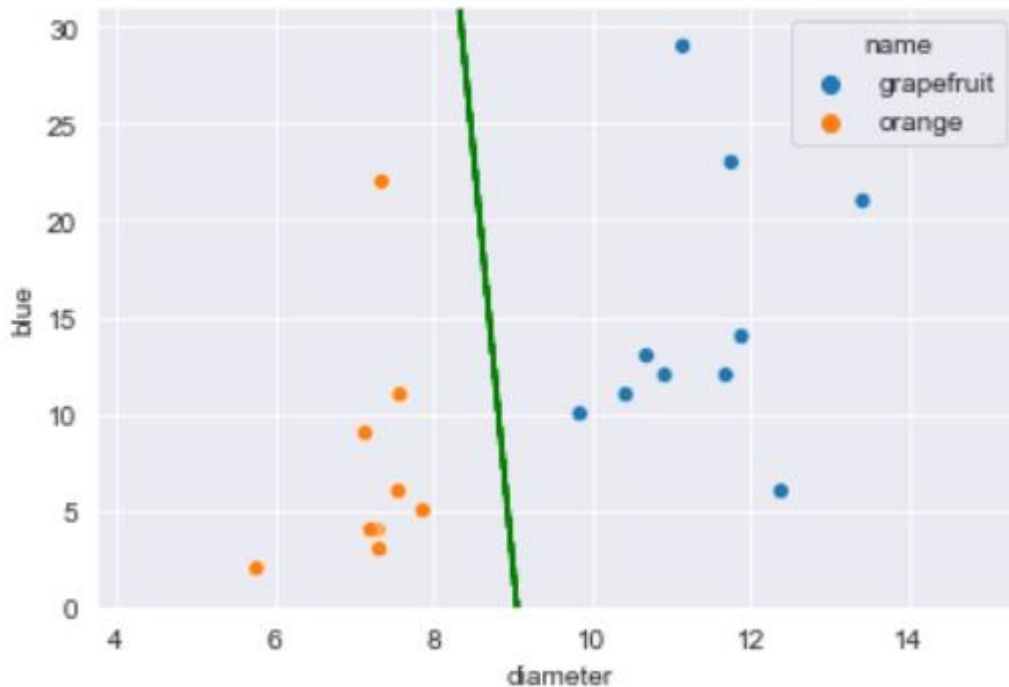
Minimalisatie dmv Gradient Descent

mis = hoge fout



LOGISTIC REGRESSION

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



LOGISTIC REGRESSION – HIGHER ORDER FEATURES

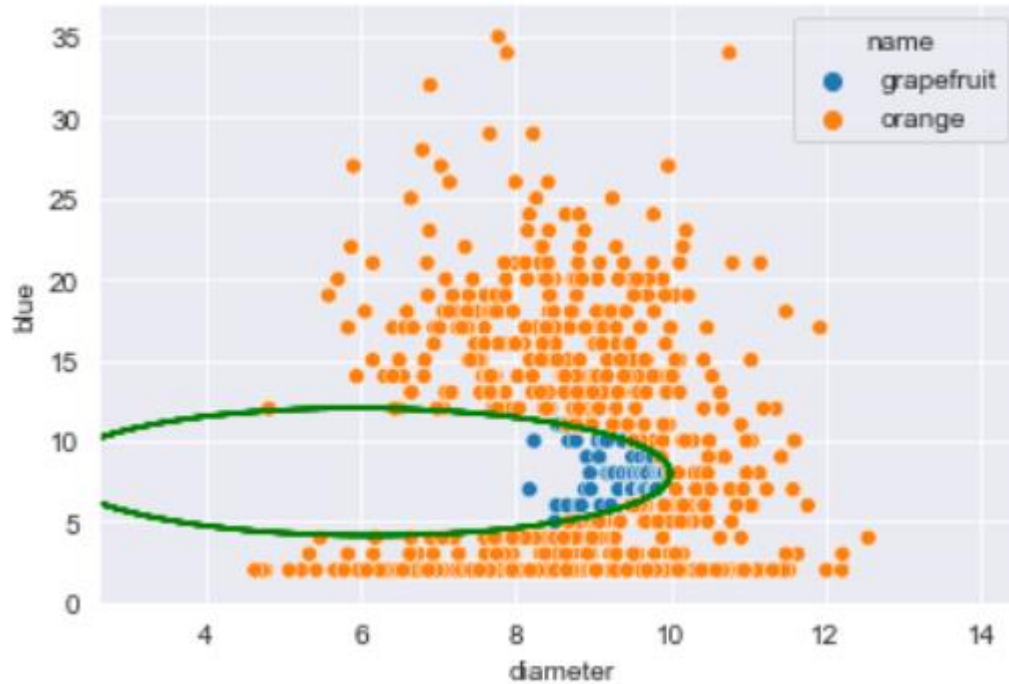


A scatter plot showing the relationship between 'diameter' (x-axis, ranging from 4 to 12) and 'blue' (y-axis, ranging from 0 to 35). The data points are categorized by 'name' (grapefruit, represented by blue dots, and orange, represented by orange dots). A green line represents a linear decision boundary, separating the two classes. The plot illustrates that a simple linear model may not be sufficient for separating the two classes, as many orange points are located to the right of the decision boundary.

diameter	blue	name
5.5	13	orange
6.0	2	orange
6.2	2	orange
6.5	2	orange
6.8	2	orange
7.0	2	orange
7.2	2	orange
7.5	2	orange
7.8	2	orange
8.0	2	orange
8.2	2	orange
8.5	2	orange
8.8	2	orange
9.0	2	orange
9.2	2	orange
9.5	2	orange
9.8	2	orange
10.0	2	orange
10.2	2	orange
10.5	2	orange
10.8	2	orange
11.0	2	orange
11.2	2	orange
11.5	2	orange
11.8	2	orange
12.0	5	orange
6.0	17	orange
6.5	18	orange
7.0	13	orange
7.5	13	orange
8.0	16	orange
8.2	19	orange
8.5	12	orange
8.8	15	orange
9.0	18	orange
9.2	12	orange
9.5	15	orange
9.8	13	orange
10.0	14	orange
10.2	19	orange
10.5	12	orange
10.8	14	orange
11.0	14	orange
11.2	14	orange
11.5	14	orange
11.8	14	orange
12.0	5	orange
6.5	34	orange
8.0	29	orange
8.2	35	orange
10.8	34	orange
8.5	11	grapefruit
9.0	11	grapefruit
9.5	10	grapefruit
10.0	6	grapefruit
10.2	9	grapefruit
10.5	6	grapefruit
10.8	7	grapefruit
11.0	10	grapefruit
11.2	7	grapefruit
11.5	10	grapefruit
11.8	7	grapefruit
12.0	7	grapefruit



LOGISTIC REGRESSION – HIGHER ORDER FEATURES

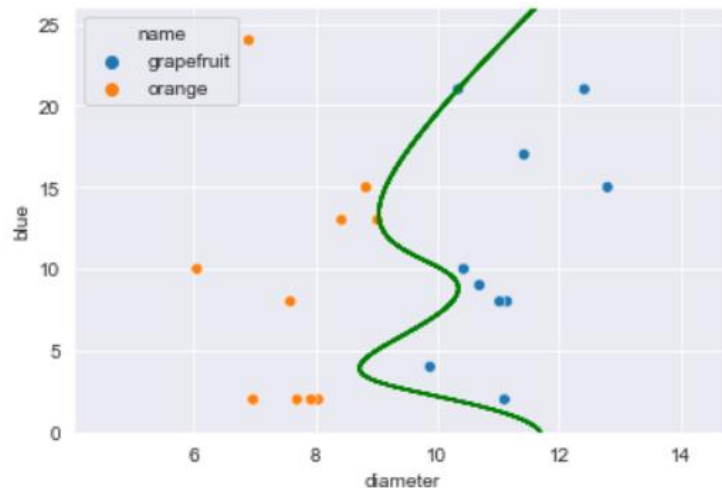


LOGISTIC REGRESSION – REGULARISATIE

Regularisatie via C-parameter

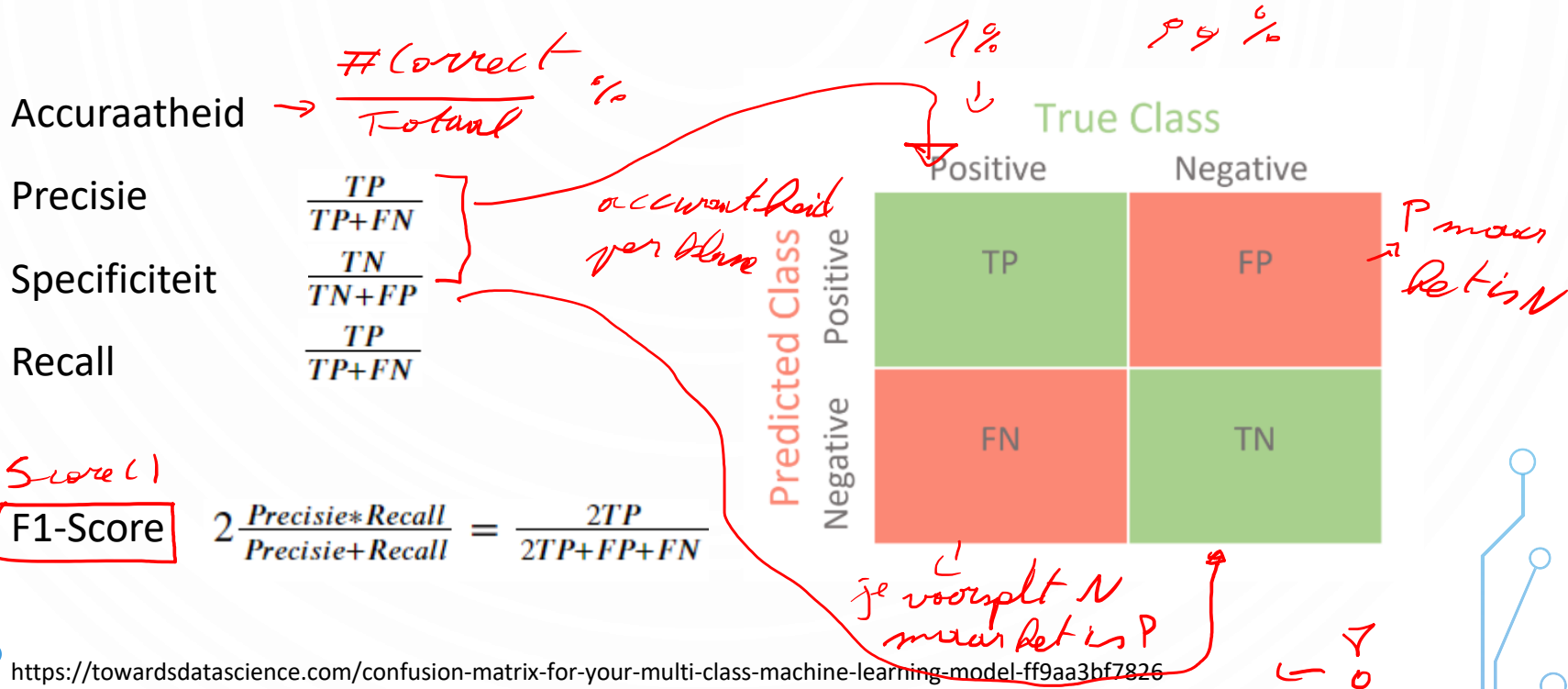
Inverse regularisatie sterkte

Hoge waarde = weinig regularisatie

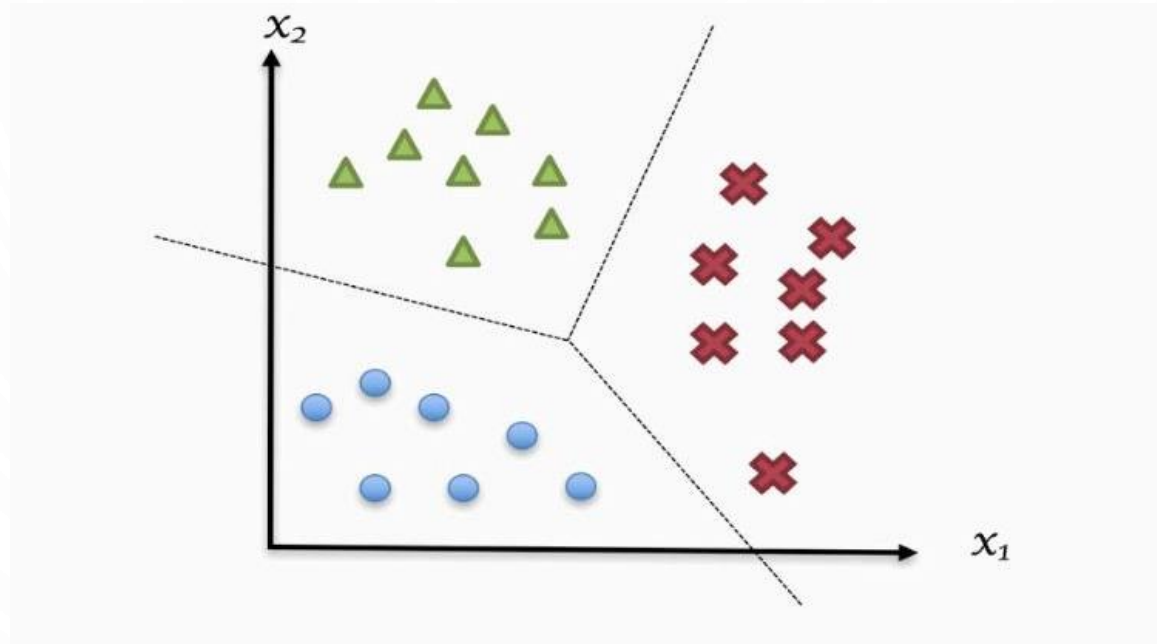


```
model = LogisticRegression(C=10) # C= inverse regularisatiesterkte  
model.fit(X, df_trimmed.output)
```

LOGISTIC REGRESSION – EVALUATIE



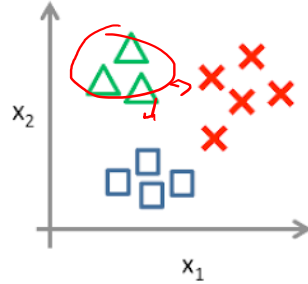
LOGISTIC REGRESSION – MULTICLASS



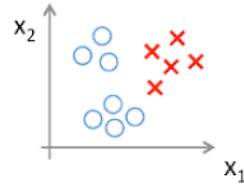
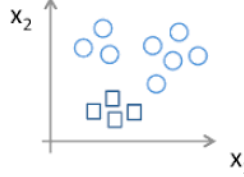
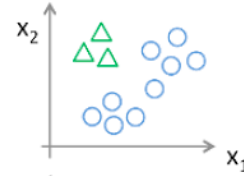
Multi-class classificatie bestaat uit meerdere

LOGISTIC REGRESSION – ONE VS ALL *binair classificatie*

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



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

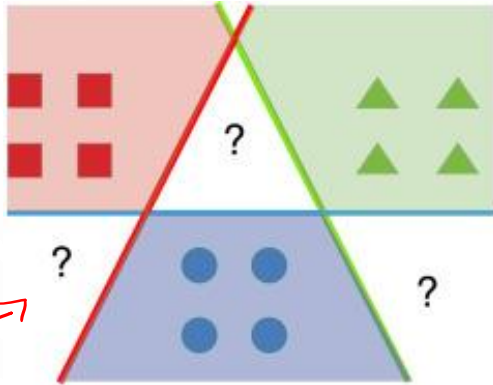


*N classificatie nodig
1/klass*

*Conflicten mogelijk
→ Kies met de
hoogste kans*

LOGISTIC REGRESSION – ONE VS ONE

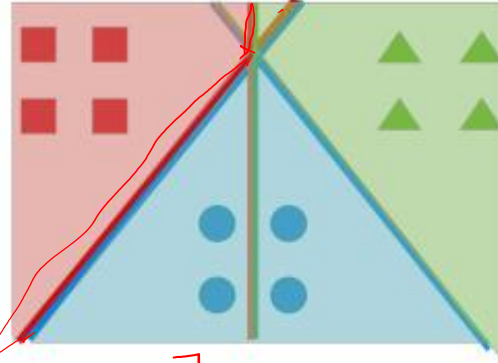
One vs all



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

		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

LOGISTIC REGRESSION – 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