Machine Learning

3. classificatie en logistische regressie

ml:classificatie

binaire classificatie

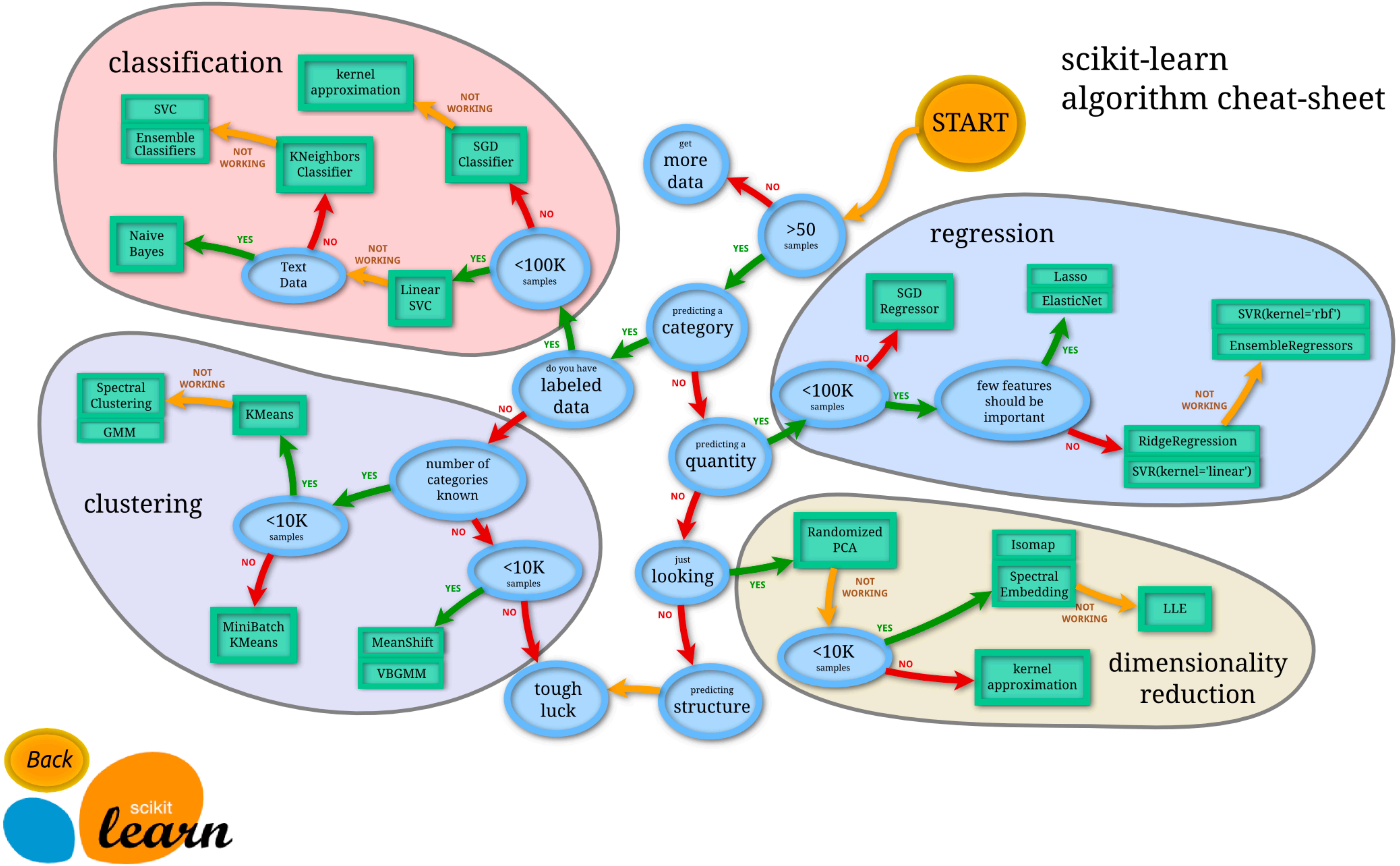
- kanker kwaadaardig?
- e-mail spam?
- online transactie betrouwbaar?
- bloemwilg virginica?
- is dit Henk Tattje?

multi-class classificatie

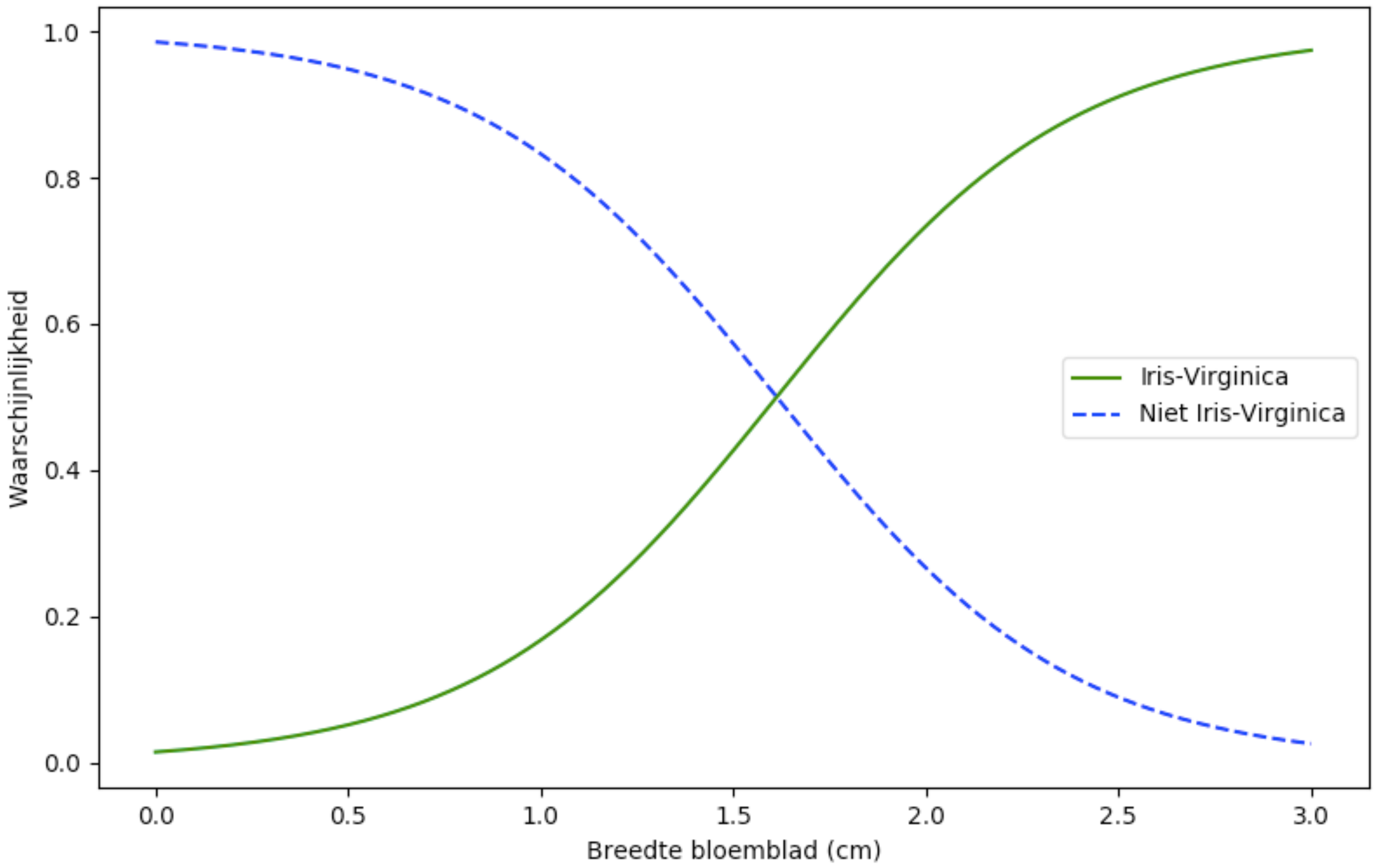
- wat voor soort insect?
- welk cijfer staat op dit plaatje?
- wie staat er op deze foto?
- wat voor soort muziekstuk is dit?

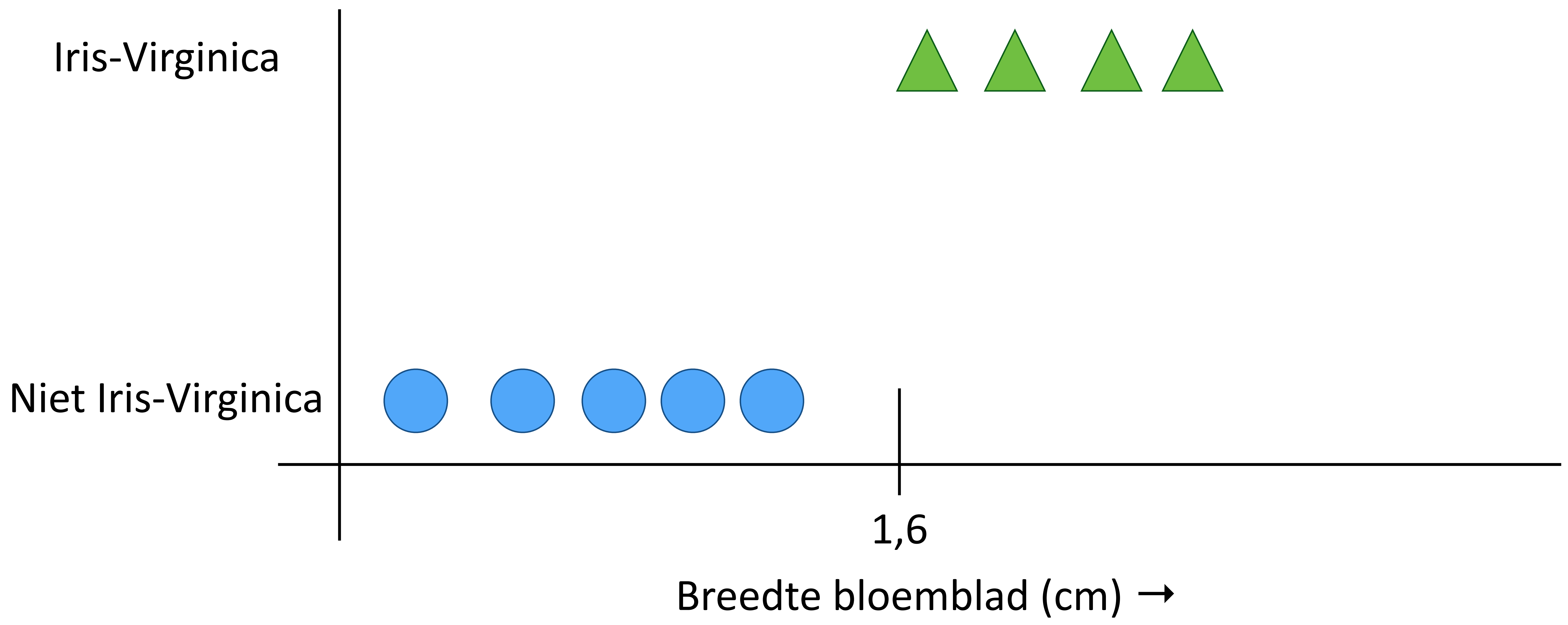
classificatie van classificatie-algoritmen

- Linear classifiers
  - Fisher's linear discriminant
  - Logistic regression
  - Naive Bayes classifier
  - Perceptron
- Support vector machines
  - Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
  - k-nearest neighbor
- Boosting (meta-algorithm)
- Decision trees
  - Random forests
- Neural networks
- Learning vector quantization


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
NO
YES
predicting a category
do you have labeled data
predicting a quantity
just looking
predicting structure
tough luck
classification
kernel approximation
SVC
Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K samples
NOT WORKING
regression
SGD Regressor
Lasso
ElasticNet
SVR(kernel='rbf')
EnsembleRegressors
<100K samples
few features should be important
RidgeRegression
SVR(kernel='linear')
clustering
Spectral Clustering
GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
MeanShift
VBGMM
dimensionality reduction
Randomized PCA
Isomap
Spectral Embedding
LLE
<10K samples
kernel approximation
Back
scikit learn

https://en.wikipedia.org/wiki/Iris_flower_data_set


Waarschijnlijkheid
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Breedte bloemblad (cm)
Iris-Virginica
Niet Iris-Virginica


Iris-Virginica
Niet Iris-Virginica
1,6
Breedte bloemblad (cm) →

Binaire classificatie

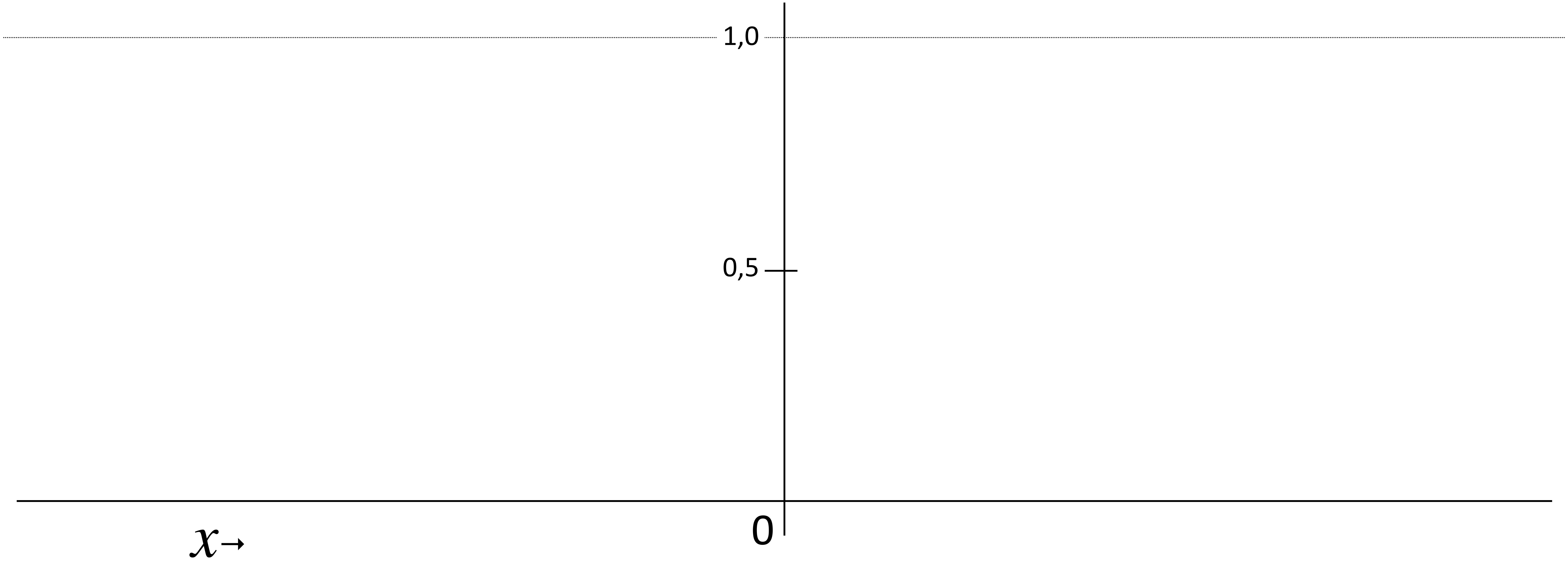
$$y \in \{0,1\}$$

$$0 \leq h_\theta(x) \leq 1$$

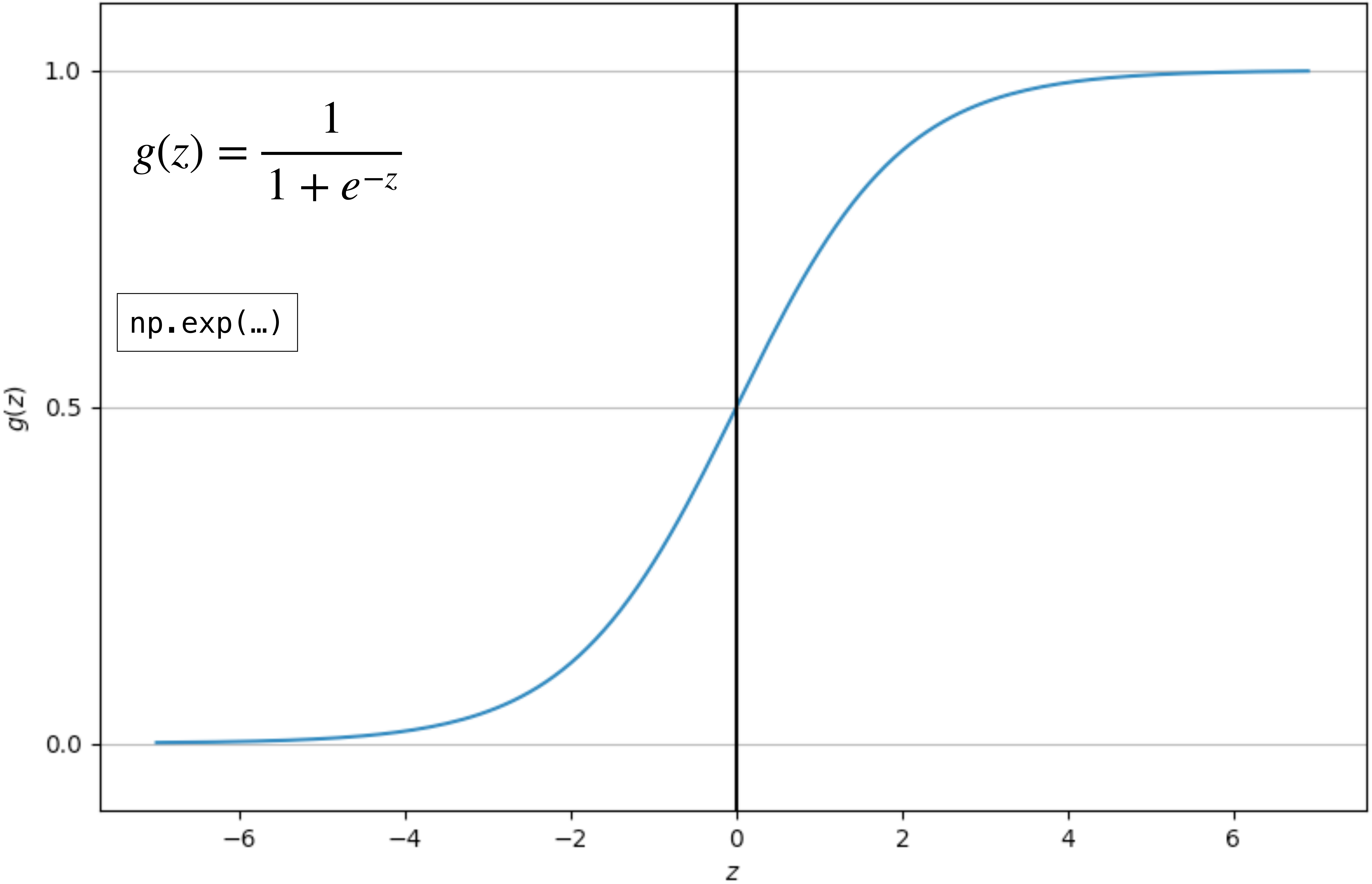
$$\boxed{\begin{array}{c}\text{Lineaire regressie}\\ h_\theta(x) = \theta^T x\end{array}}$$

logistische regressie

$$0 \leq h_\theta(x) \leq 1$$



sigmoïdefunctie

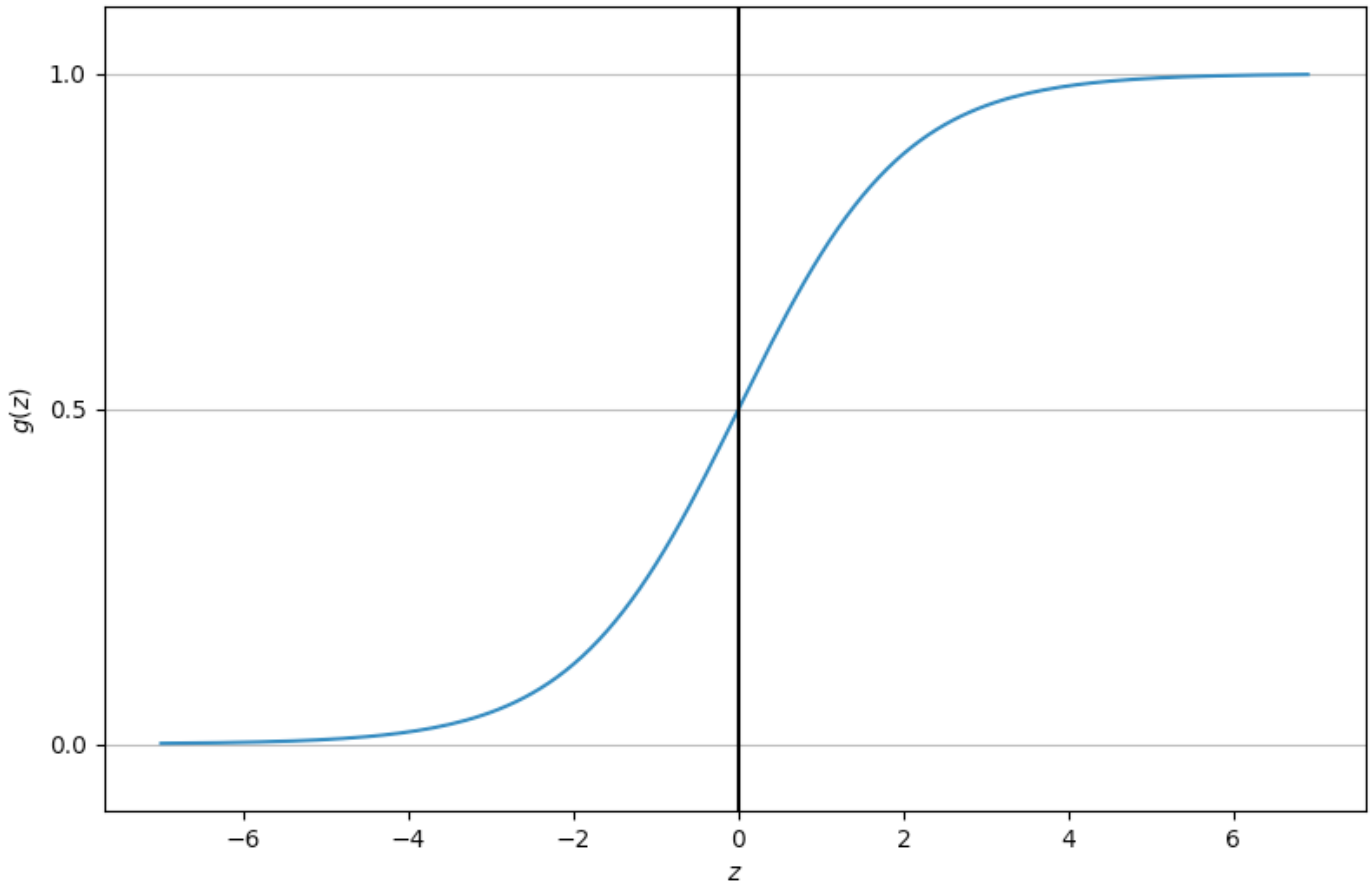


Wat betekent die voorspelling?

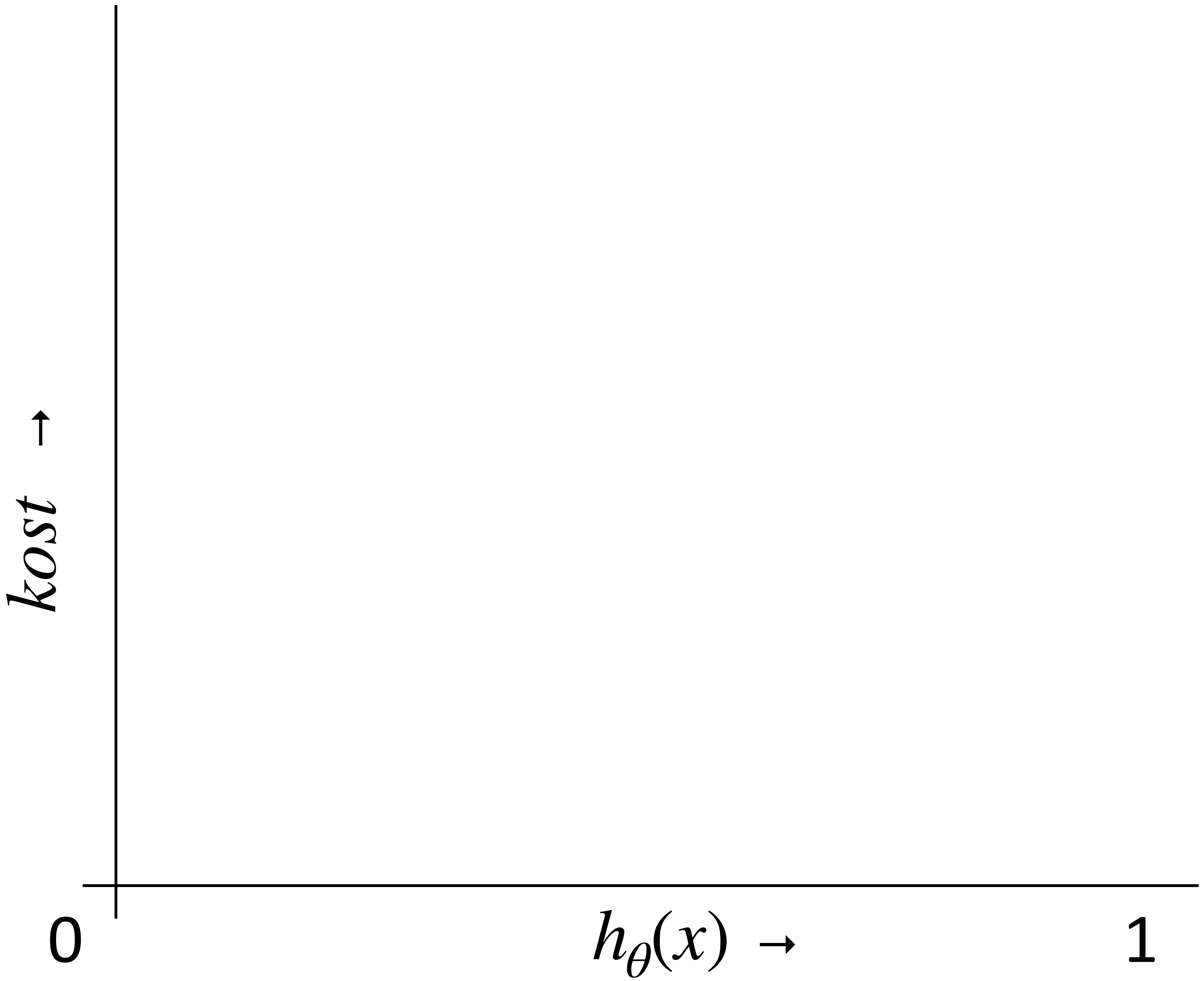
$$X = \begin{bmatrix} 1 & 1{,}6 \\ 1 & 2{,}5 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 0{,}3 \\ 0{,}125 \end{bmatrix}$$

ml:classificatie kostenfunctie

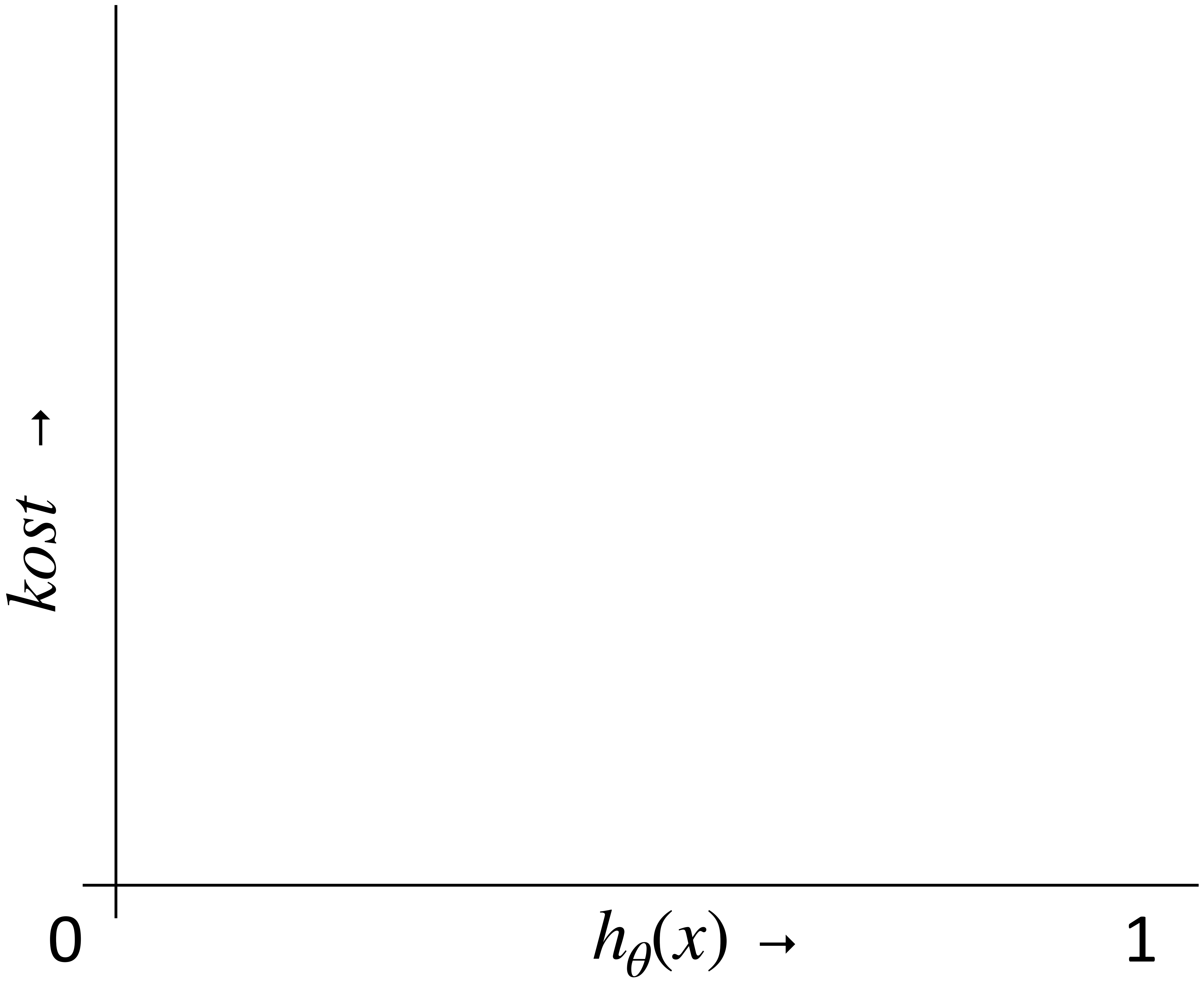

1.0
0.5
0.0
g(z)
−6
−4
−2
0
2
4
6
z

$$Kost = \begin{cases} -log(h_\theta(x)) & \text{als } y = 1 \\ -log(1 - h_\theta(x)) & \text{als } y = 0 \end{cases}$$



$y = 1$

$$Kost = \begin{cases} -log(h_\theta(x)) & \text{als } y = 1 \\ -log(1 - h_\theta(x)) & \text{als } y = 0 \end{cases}$$



$y = 0$

$$Kost = \begin{cases} -log(h_\theta(x)) & \text{als } y = 1 \\ -log(1 - h_\theta(x)) & \text{als } y = 0 \end{cases}$$

$$= -\, ylog(h_\theta(x)) - (1 - y)log(1 - h_\theta(x))$$

$$= ylog(h_\theta(x)) + (1 - y)log(1 - h_\theta(x))$$

Kostenfunctie voor logistische regressie

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}log(h_\theta(x^{(i)})) + (1 - y^{(i)})log(1 - h_\theta(x^{(i)}))\right]$$

Dat ziet er indrukwekkend uit!

Maar de partiële afgeleide ervan ziet er juist weer heel vertrouwd uit:

$$\frac{\partial}{\partial\theta_j}J(\theta) = (h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

Gradient Descent bij logistische regressie

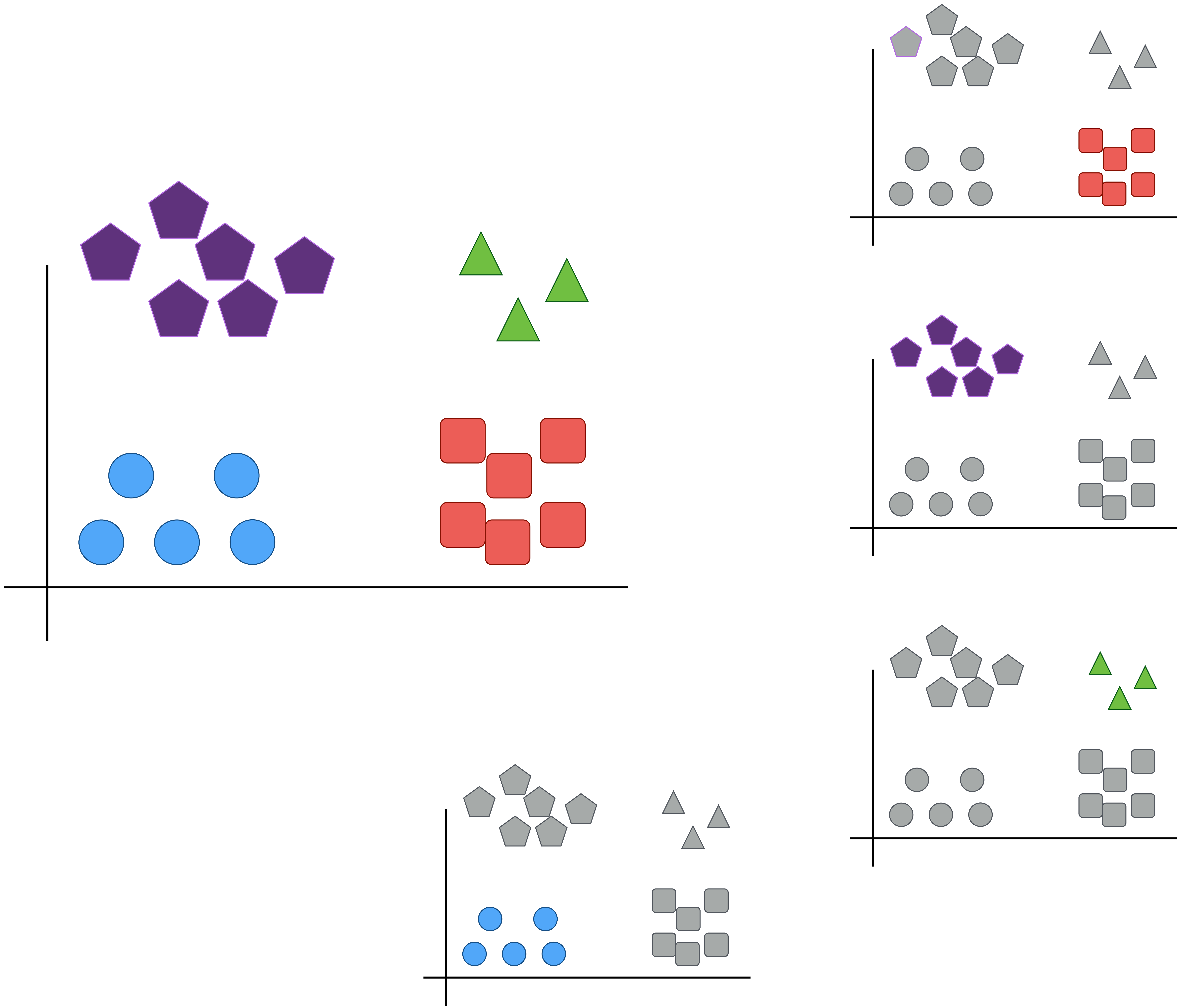
update alle $\theta_j, j = 1, j = 2, \ldots, j = n$

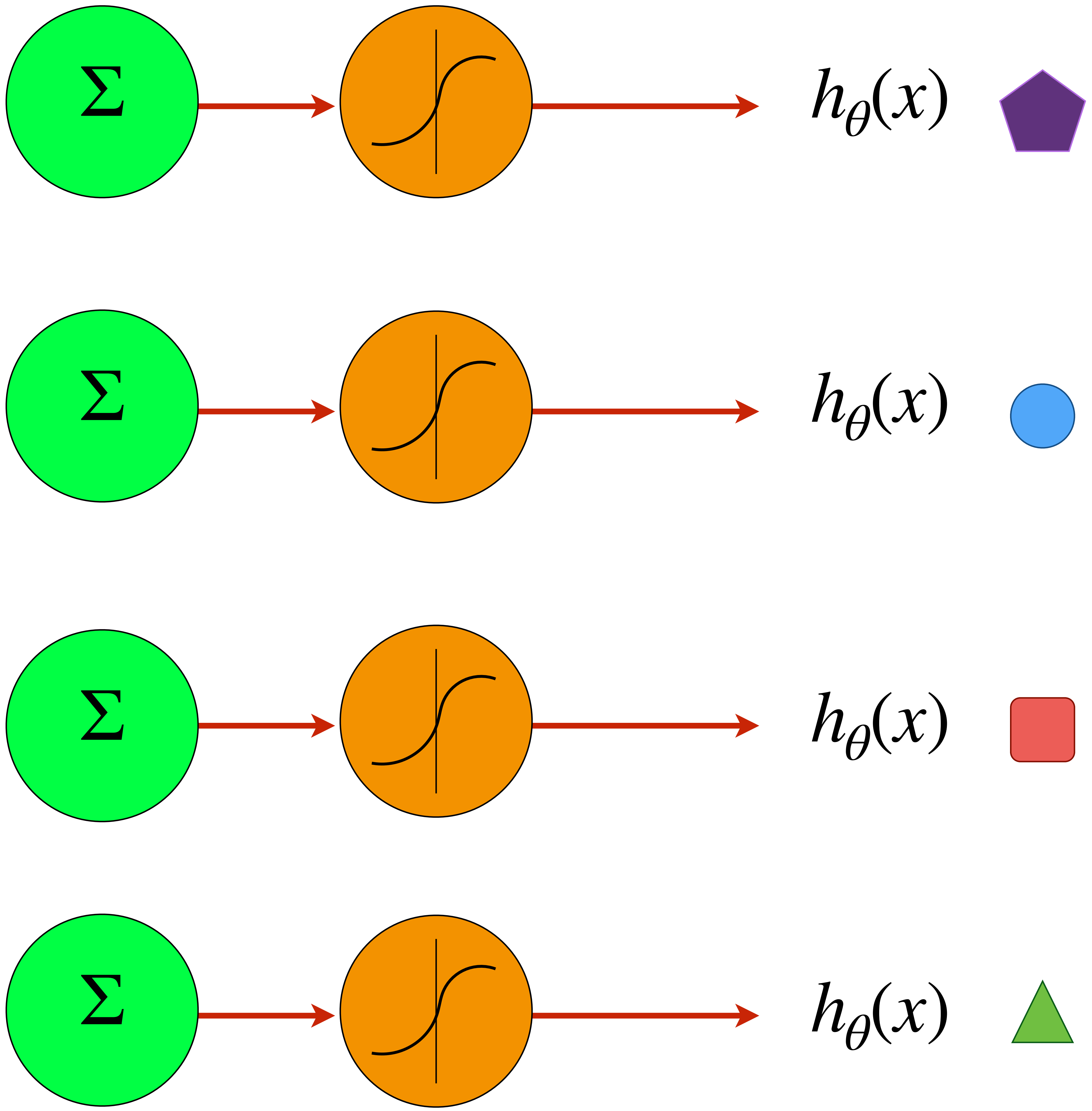
herhaal totdat een minimum bereikt is:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$:= \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

ml:multiclass classification





$$\begin{bmatrix} 0.213 \\ 0.423 \\ 0.786 \\ 0.143 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$