Underlying Blocks behind classification scores

A quick guide to machine learning

SAGEX, 30th of July 2019

Overview

- Regression
  - Description
  - Linear solution
  - Loss function

- Classification
  - Description
  - Logistic solution
  - Loss function

- Interpretation
  - Scores are probabilities ?
  - Limits of the analogy

Regression: description

Merriam-Webster dictionary: **Regression** definition:

A trend or shift toward a lower or less perfect state: such as [...]

[d] a **functional relationship** between two or more correlated variables that is often **empirically determined from data** and is used especially to predict values of one variable when given values of the others.

Use in machine learning:

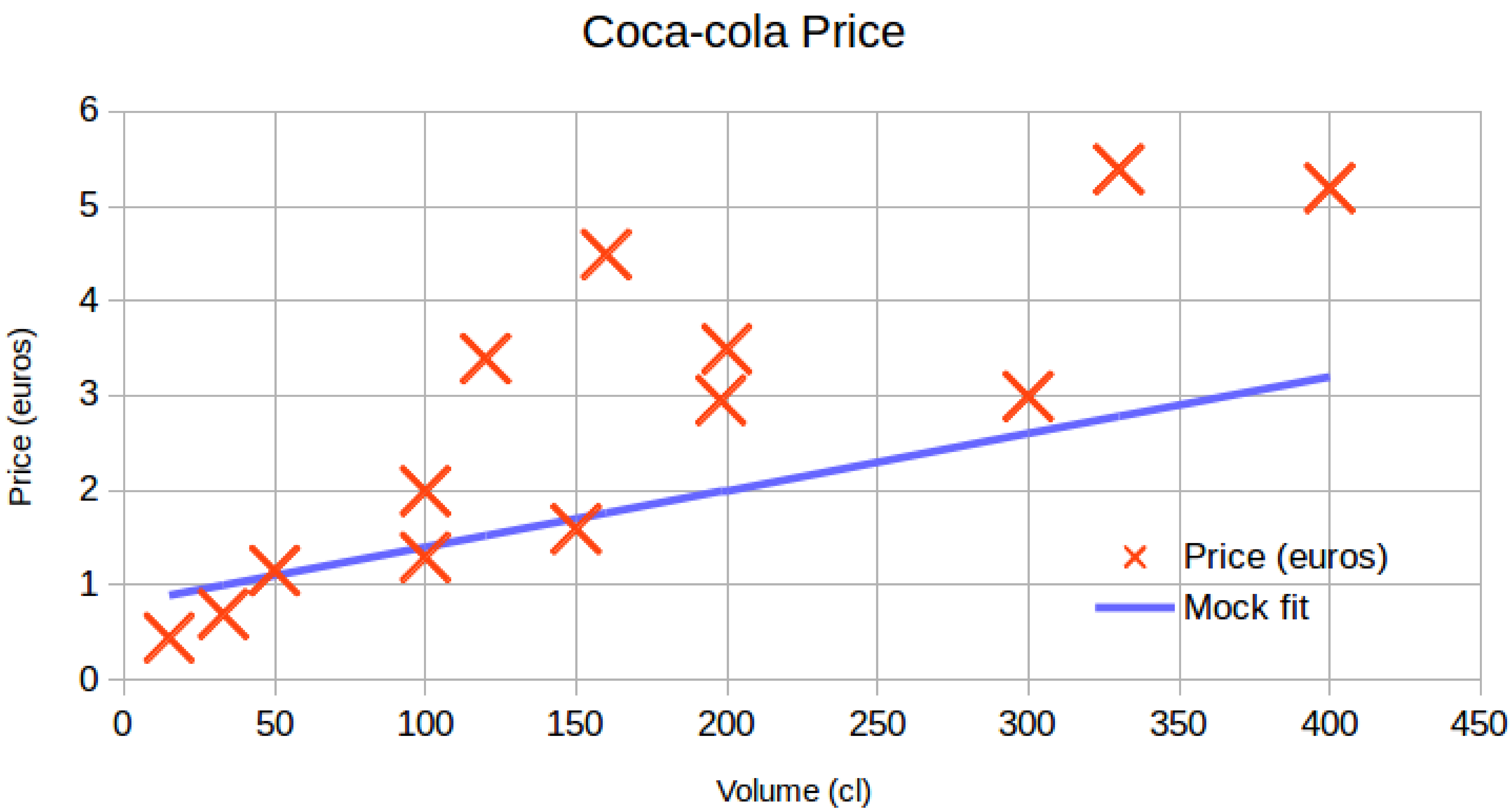
A procedure to compute numerically the **best parameters** of a given **set of functions** to **fit train-data** and the resulting method that can be **applied on test-data**.

Regression: linear solution

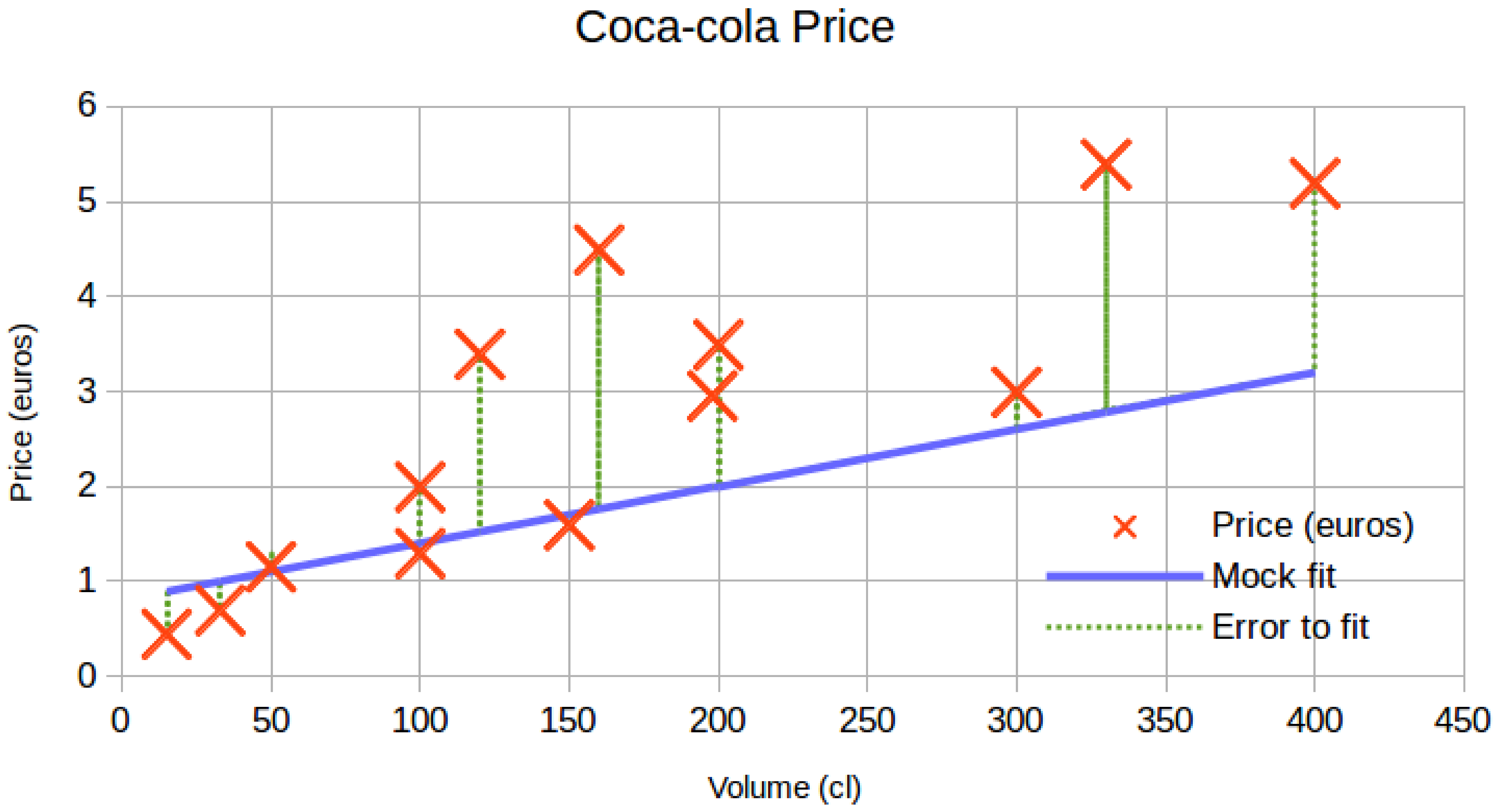
Format	Number	Volume (cl)	Price (euros)
15	1	15	0.44
33	1	33	0.69
50	1	50	1.15
25	4	100	1.99
100	1	100	1.29
15	8	120	3.39
150	1	150	1.59
20	8	160	4.49
33	6	198	2.95
50	4	200	3.49
150	2	300	2.99
33	10	330	5.39
100	4	400	5.19

https://www.monoprix.fr/search/coca-cola

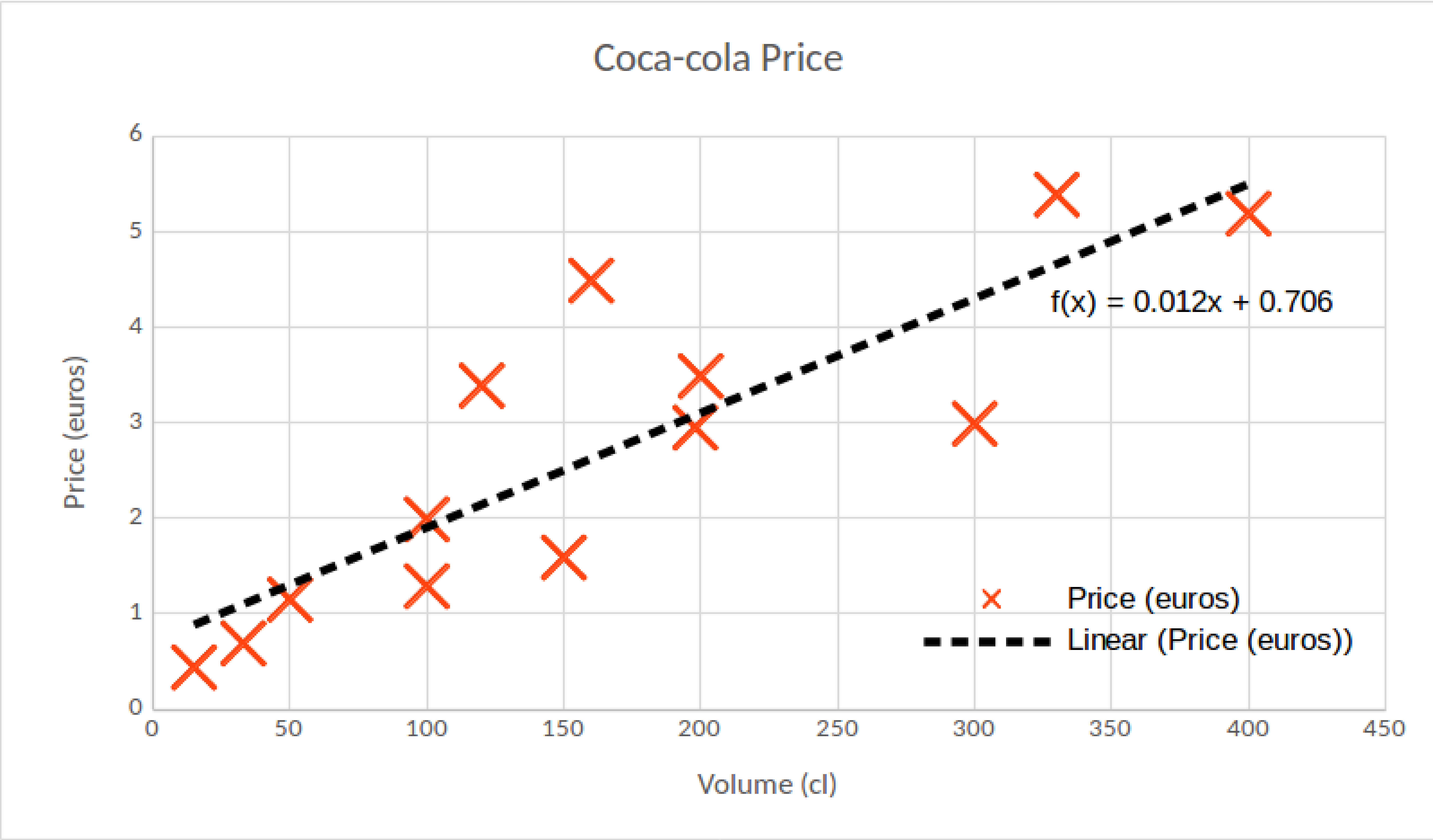
Regression: linear solution



Regression: linear solution



Regression: linear solution



Regression: loss function

Function: $$f[a,b](x) = a \times x + b$$

a and b are the parameters to optimize. x is the value of the input and f[a,b](x) is the prediction.

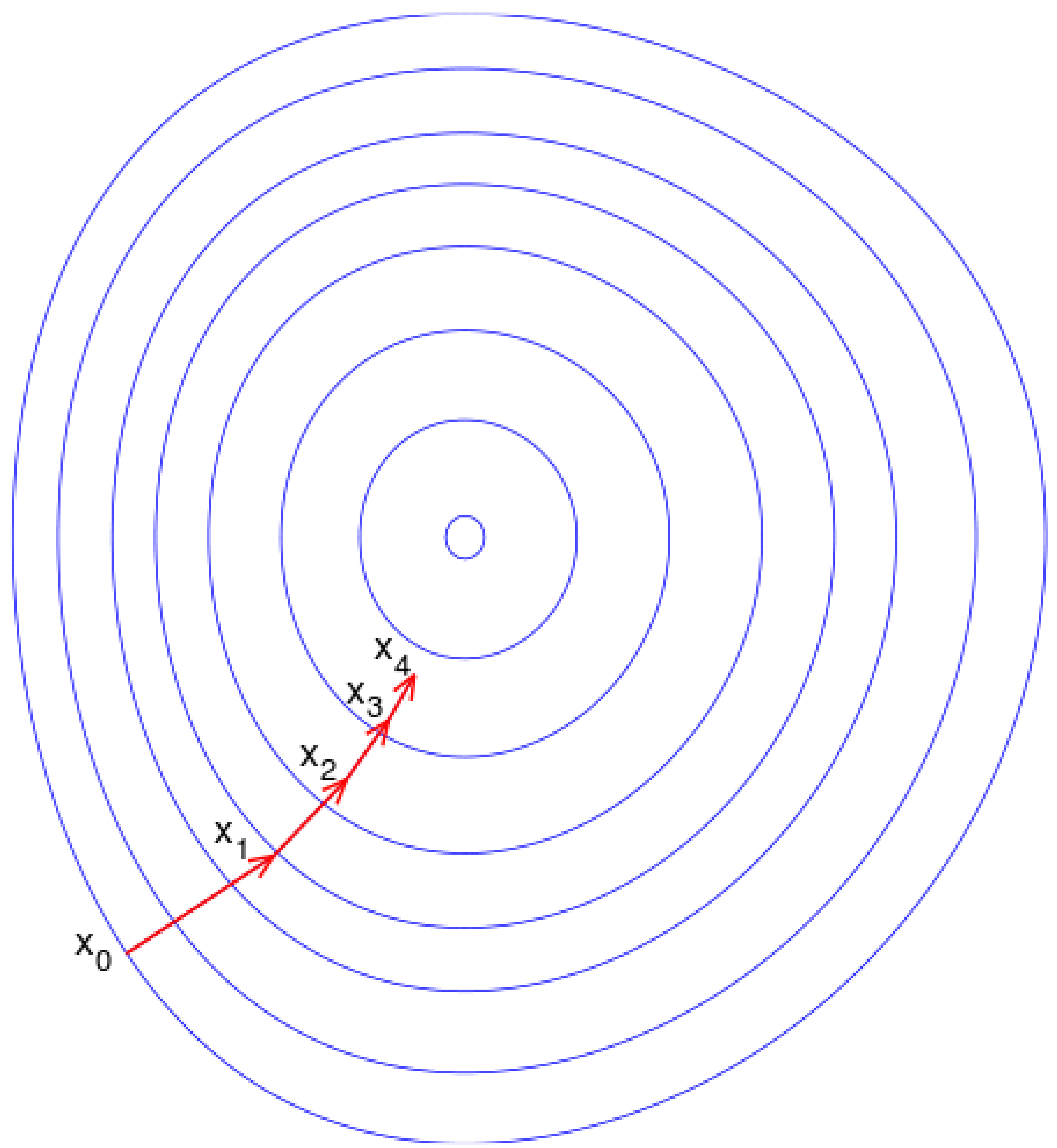
Loss/Cost function :
Mean square error

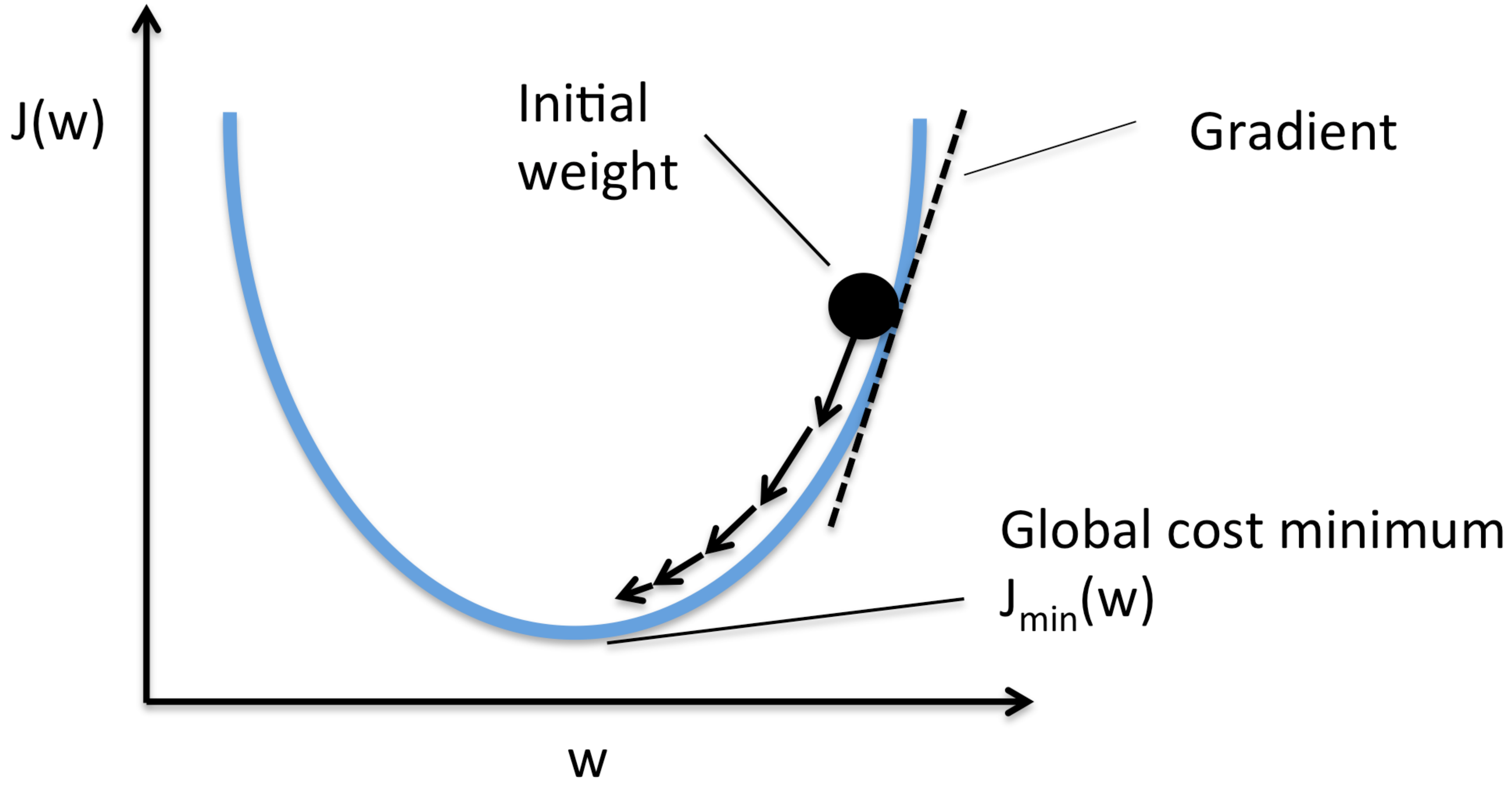
$$L(y, f[a,b]) = \sum_i (y_i - f[a,b](x_i))^2$$

Interpretation: Minimisation of the quadratic error.

Optimisation method: Gradient Descent
0) start from a given set of parameters
1) find the direction of parameters decreasing the loss
2) update the parameters according to this direction
3) go back to 1) until error is sufficiently small

Gradient descent





Classification: description

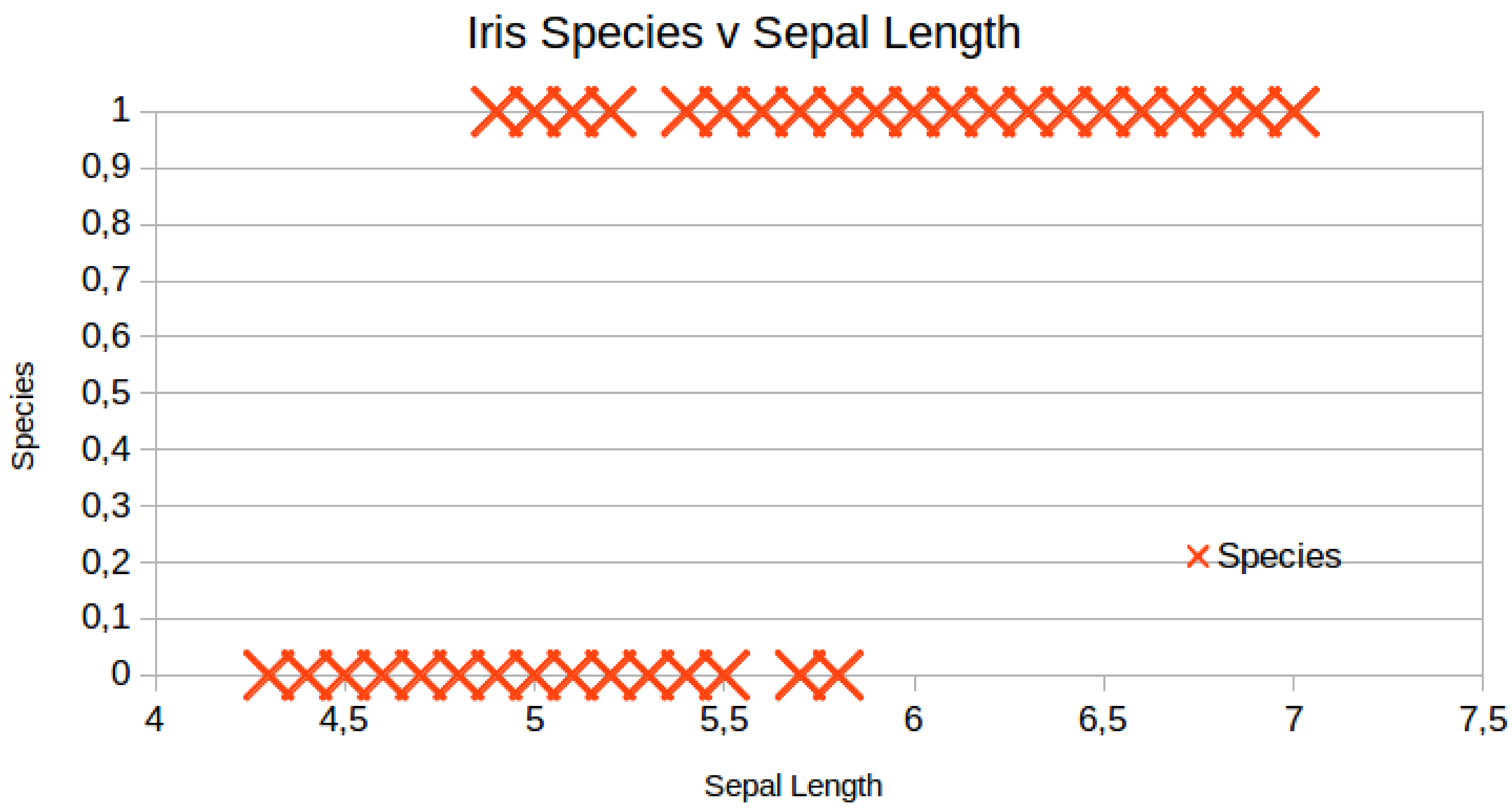
Merriam-Webster dictionary: **Classification** definition:

Systematic arrangement in groups or categories according to established criteria.

Use in machine learning:

A procedure to compute numerically the **likelihood of a element to belong to a set** using the information of train-data.

Classification: logistic solution

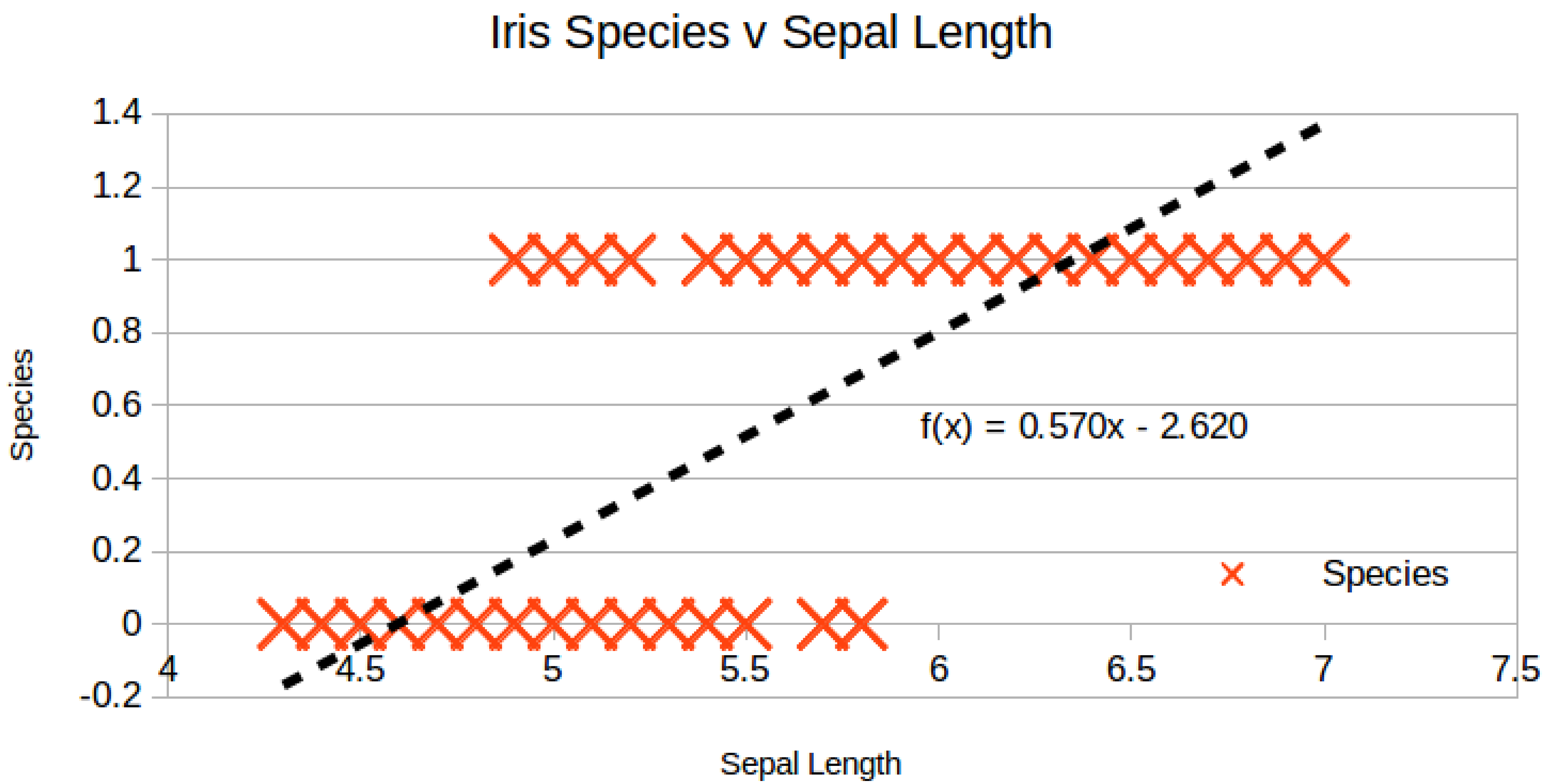


The logistic function produces continuous values between 0 and 1.

To make a predict, a threshold has to be placed in order to match the 0 or 1 classification.

https://www.kaggle.com/uciml/iris

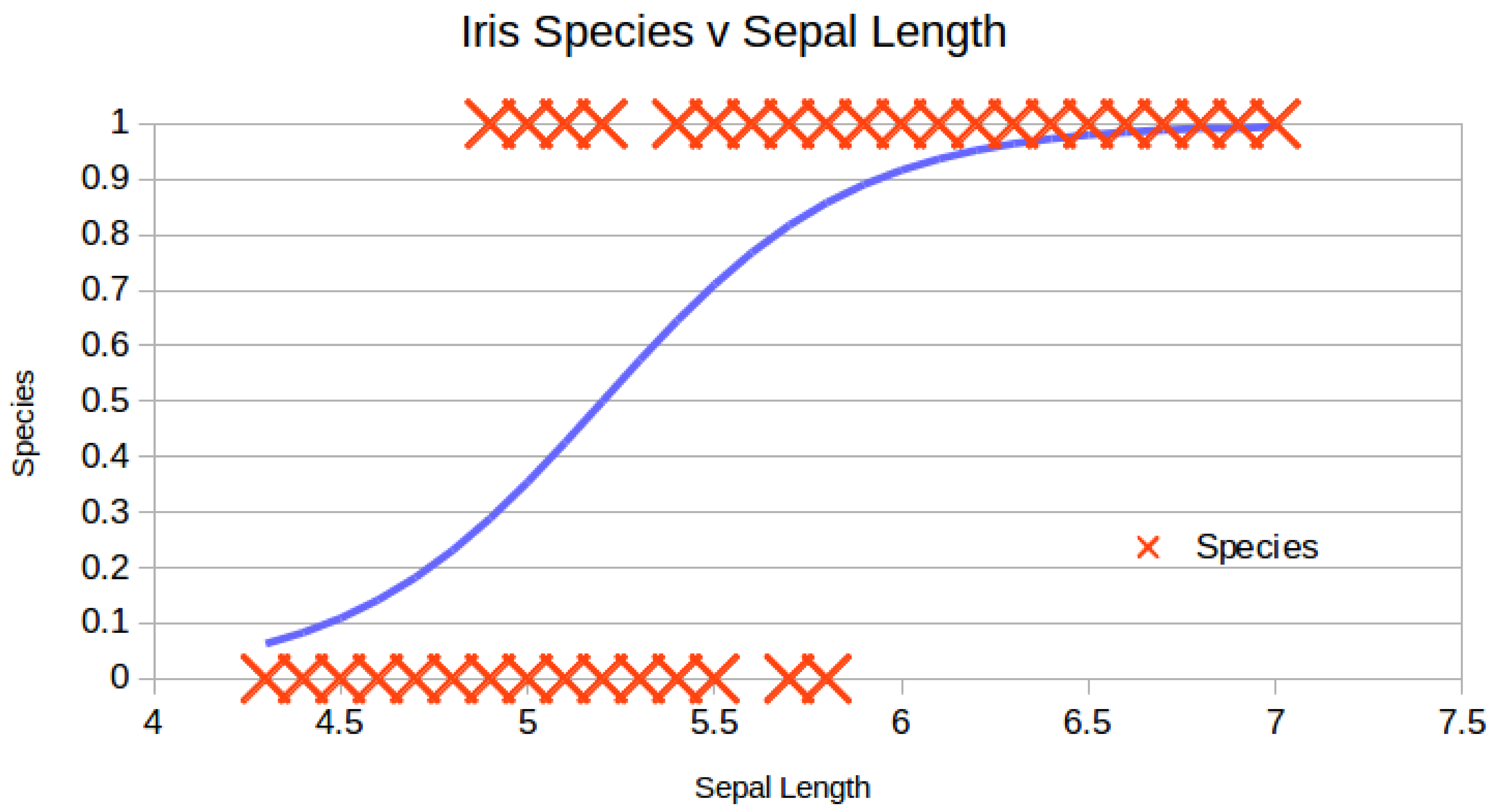
Classification: logistic solution



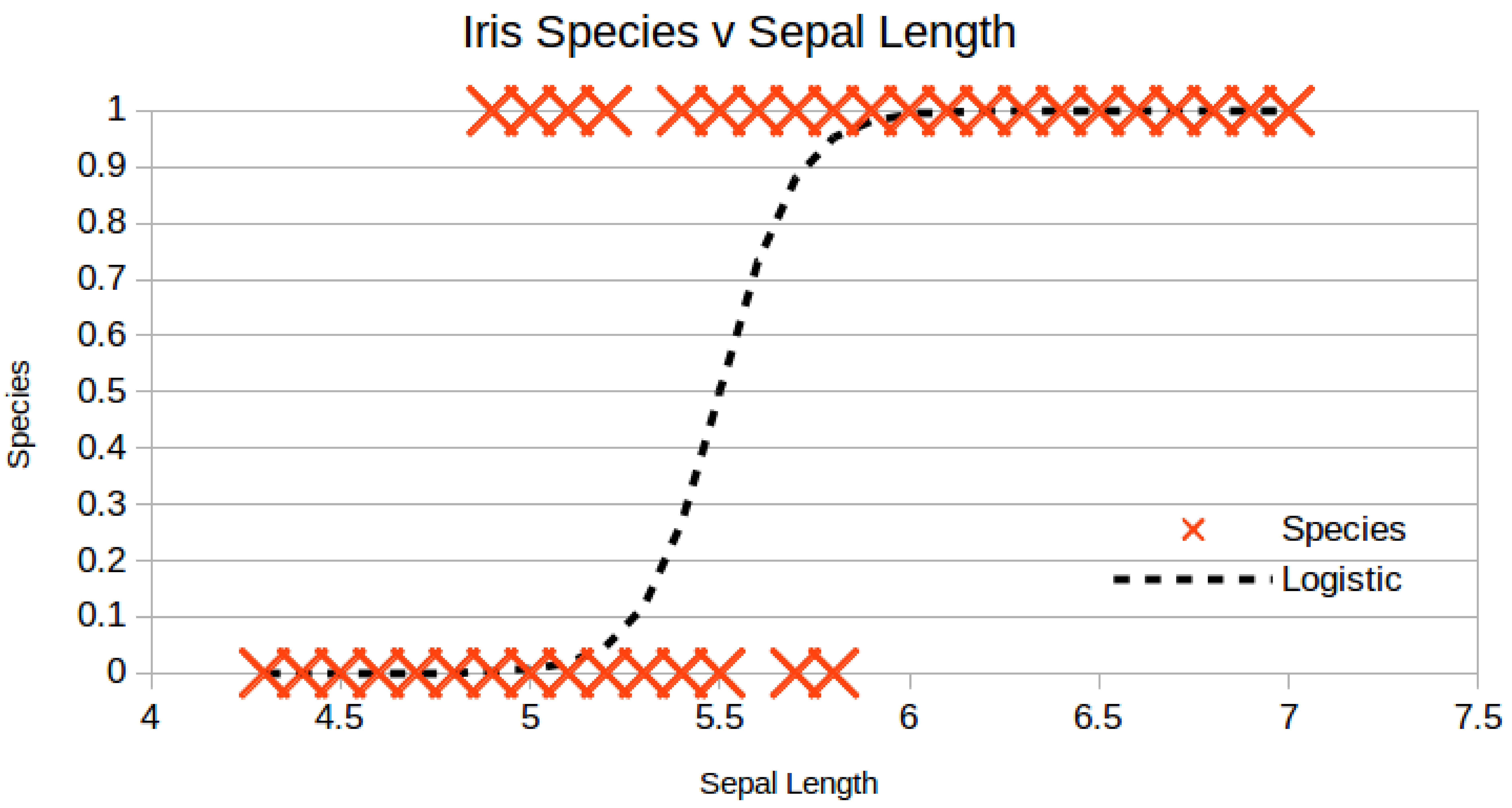
Linear regressions are not suited for classification problems because the values of the prediction function are not bounded.

The mean square error does not go to zero for high values of the input.

Classification: logistic solution



Classification: logistic solution



Classification: loss function

Function: $$f[a,b](x) = \frac{1}{1+\exp(-(a\times x)+b)}$$

a and b are the parameters to optimize. x is the value of the input and f[a,b](x) is the prediction.

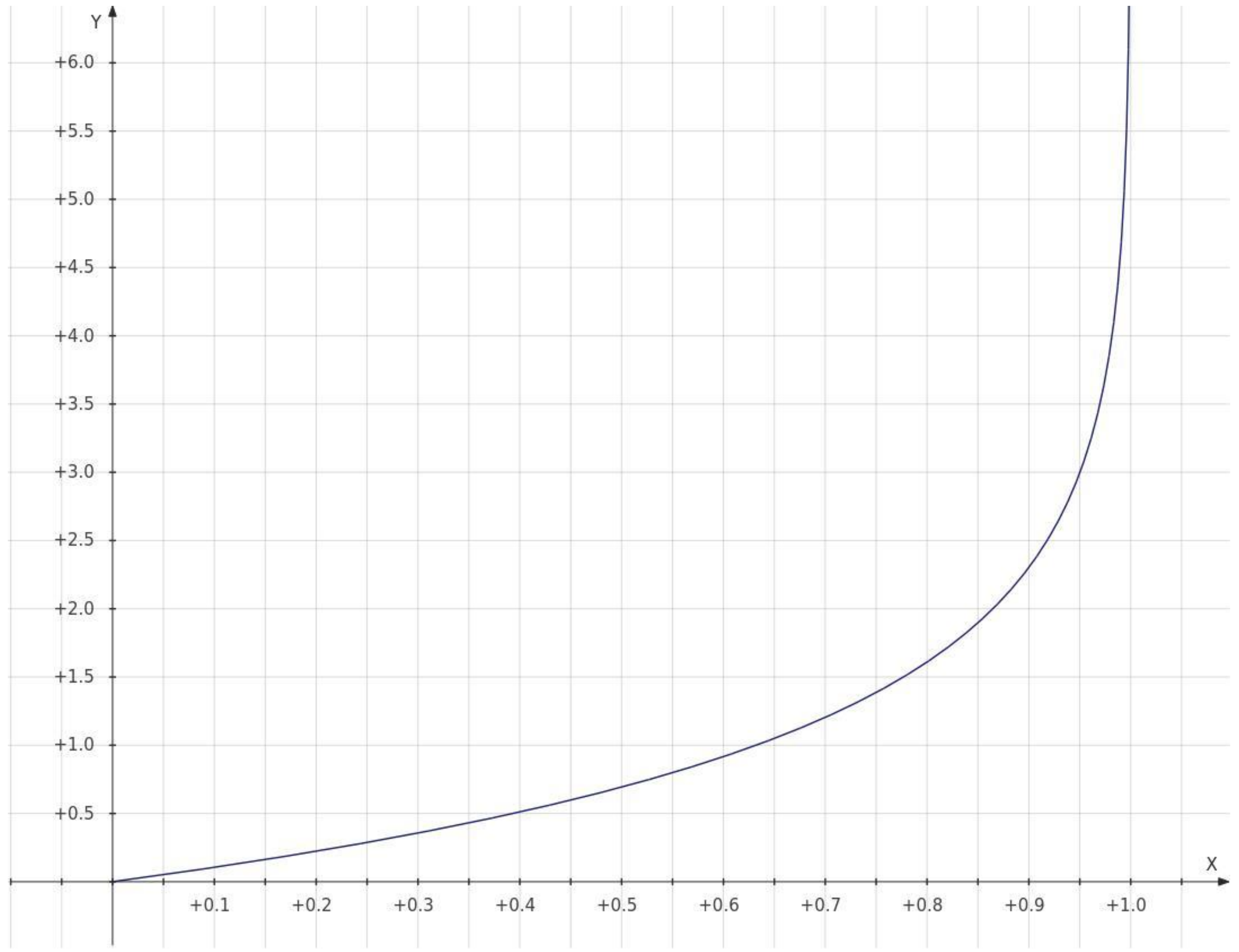
Loss/Cost function :
Binary cross-entropy

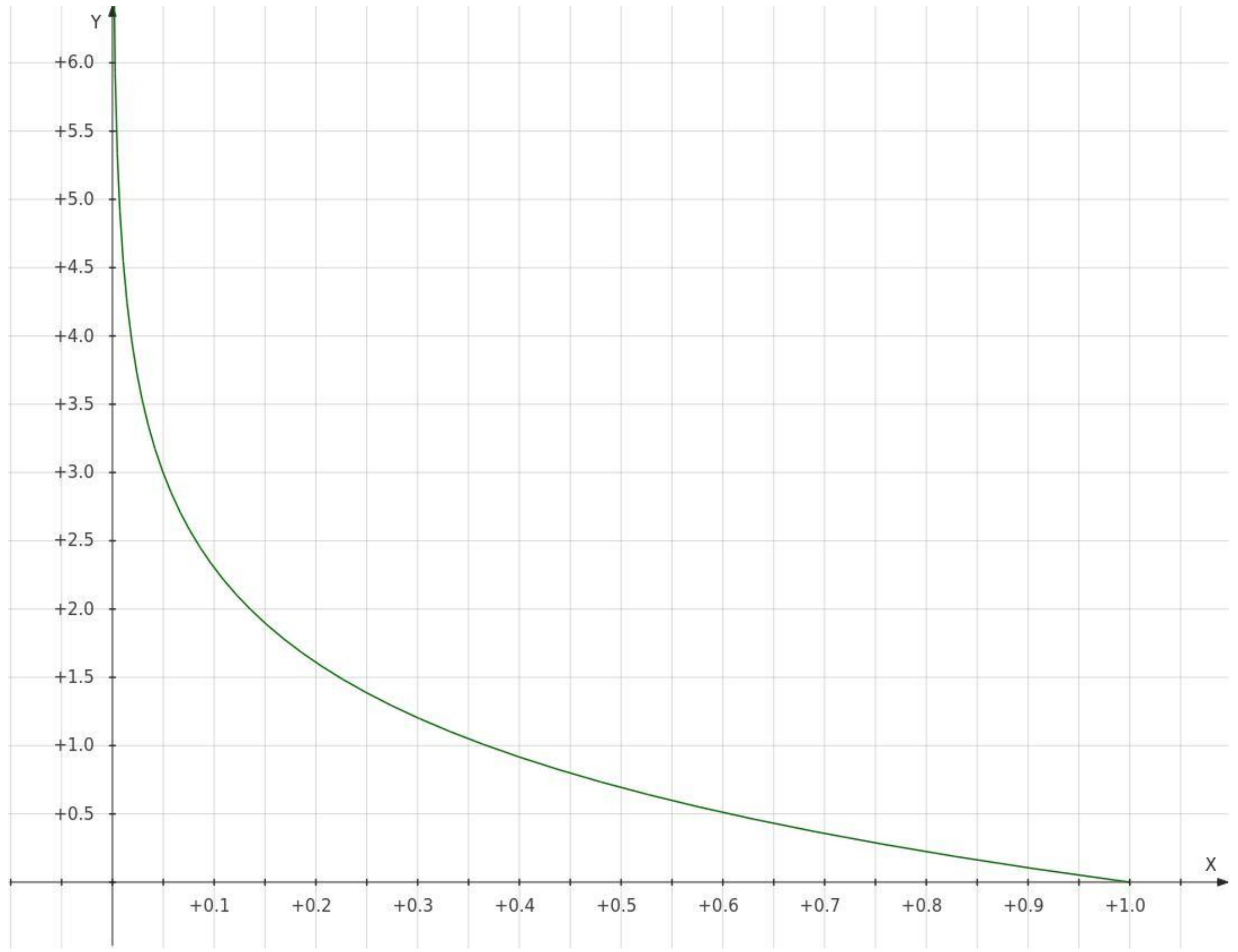
$$L(y, f[a,b]) = \sum_i \left(-y_i \ln(f[a,b](x_i)) - (1-y_i)\ln(1-f[a,b](x_i))\right)$$

Interpretation:

$y_i = 0$	$y_i = 1$
$-\ln(1-f[a,b](x_i))$	$-\ln(f[a,b](x_i))$

The loss function decreases as the logistic function gets closer to the data.

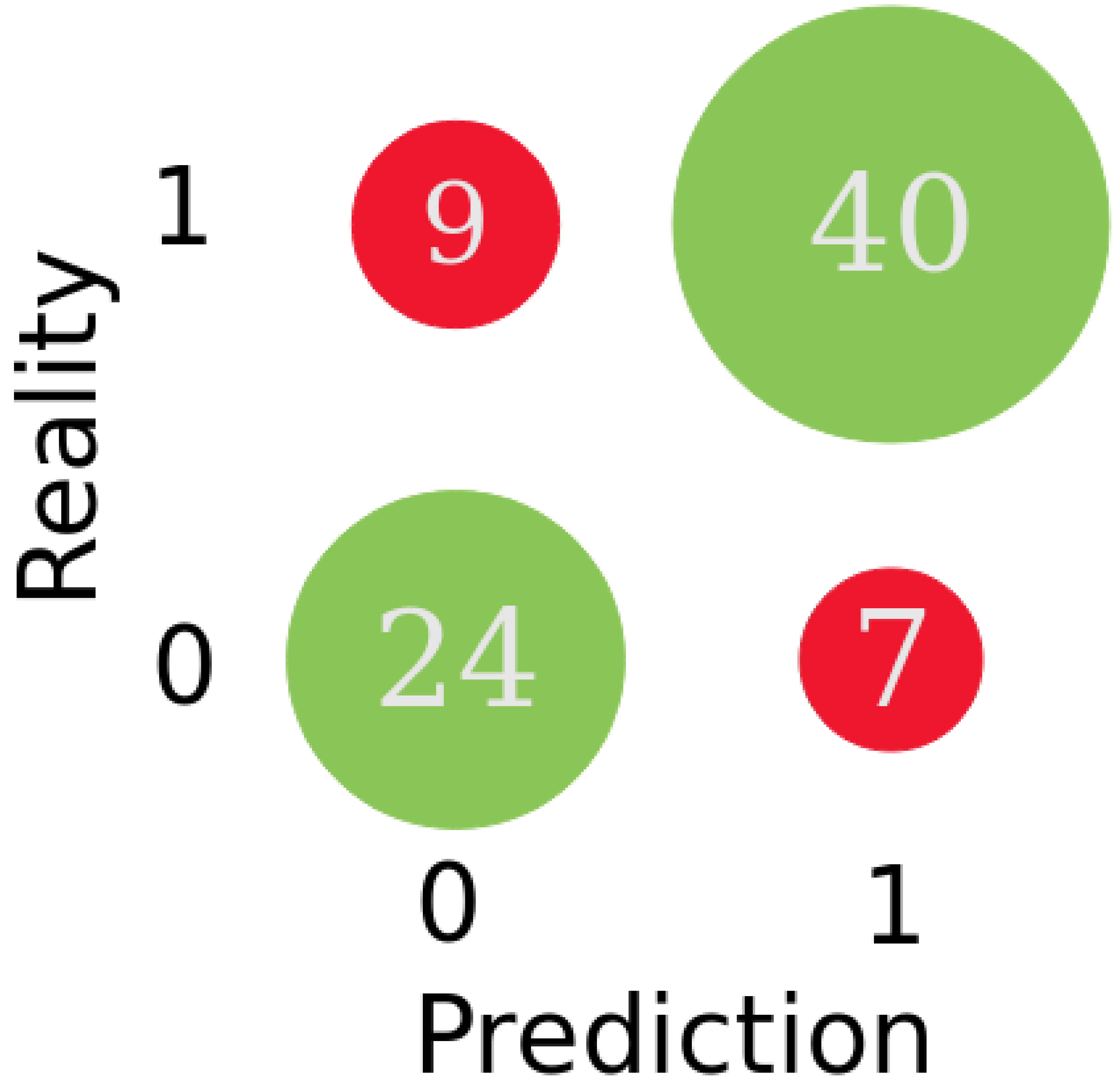




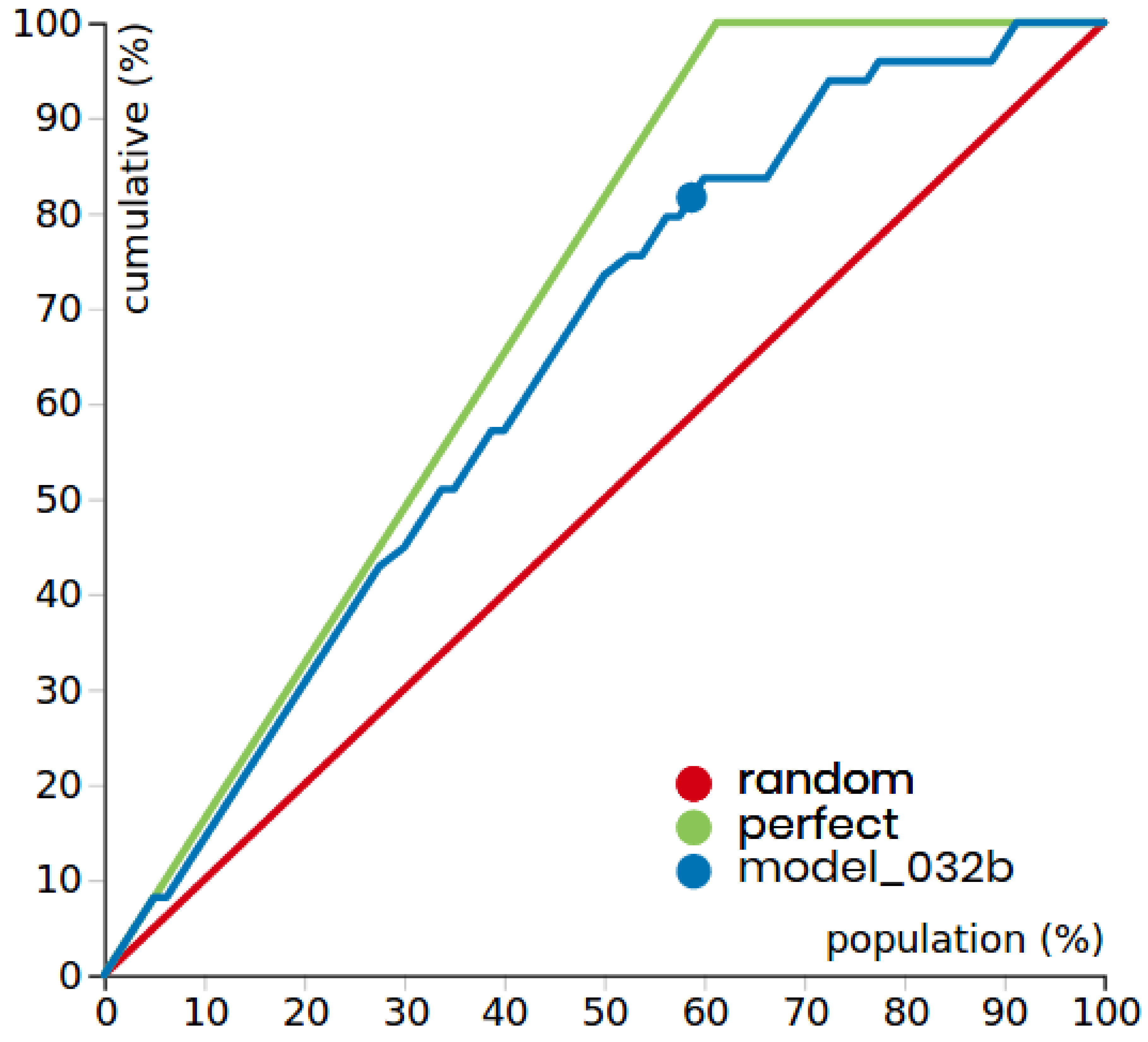
Optimisation method: Gradient Descent (identical as the regression)

Classification: success metric

Confusion matrix



Cumulative Gain Chart (Lift curve)



We use the Area Under the Curve (AUC) as a measurement of the performance of the classification:
AUC = 1 : perfect model , AUC=0.5 : random model

Scores are probabilities ?

The score is the result of the function of the model after the optimization of its parameters with respect to the loss using the train data.

The score outputted by a neural networks as properties similar to probabilities:

- Score in [0,1]
- Score of not A = 1 – Score of A
- The "likelihood" of a positive outcome increases with the score

Limits of the analogy

However, scores cannot totally be considered as a probability because:

- loss function should be adapted to the problem (done in Brain)
- predictions require to choose a threshold matching business rules
- scores of two different models cannot be compared (no Bayes' theorem).

Score ≠ Probability

THANK YOU

Alexandre CAMERON
alexandre.cameron@dreamquark.com
Jérémie ABITEBOUL
jeremie.abiteboul@dreamquark.com