Machine Learning –p2

DAVE3625

Umair Mehmood Imam

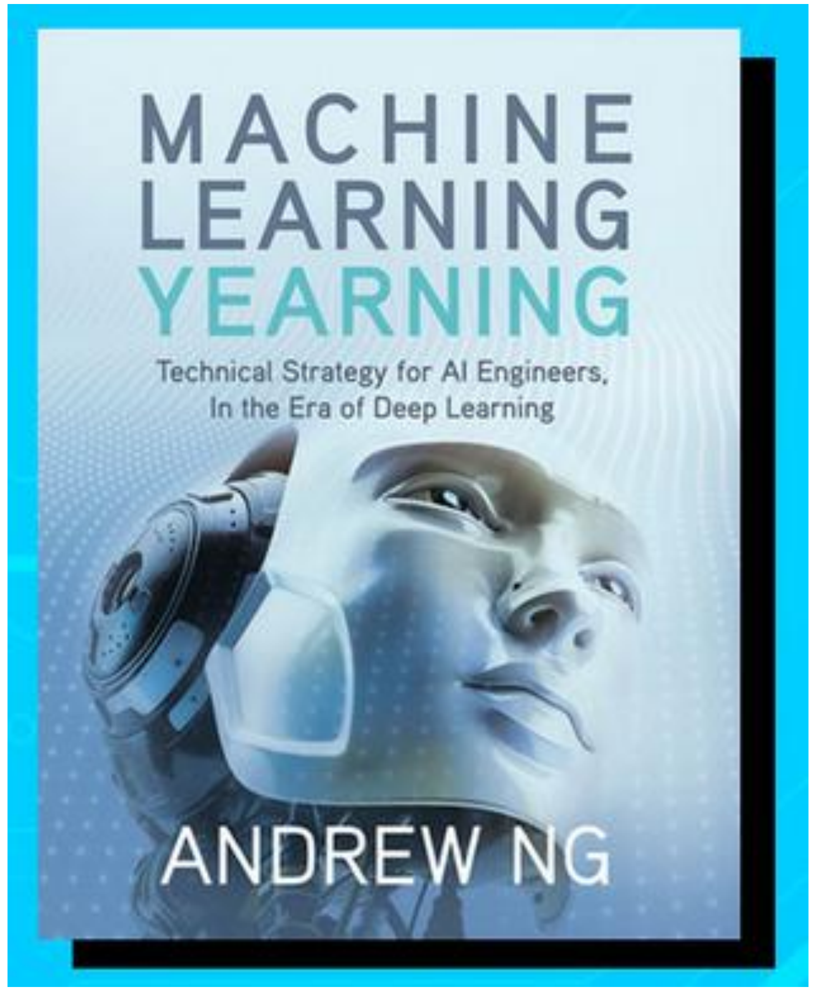
DAVE3625- INTRO TO A.I. BY UMAIR M.I

Types of Machine Learning algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Recommender systems

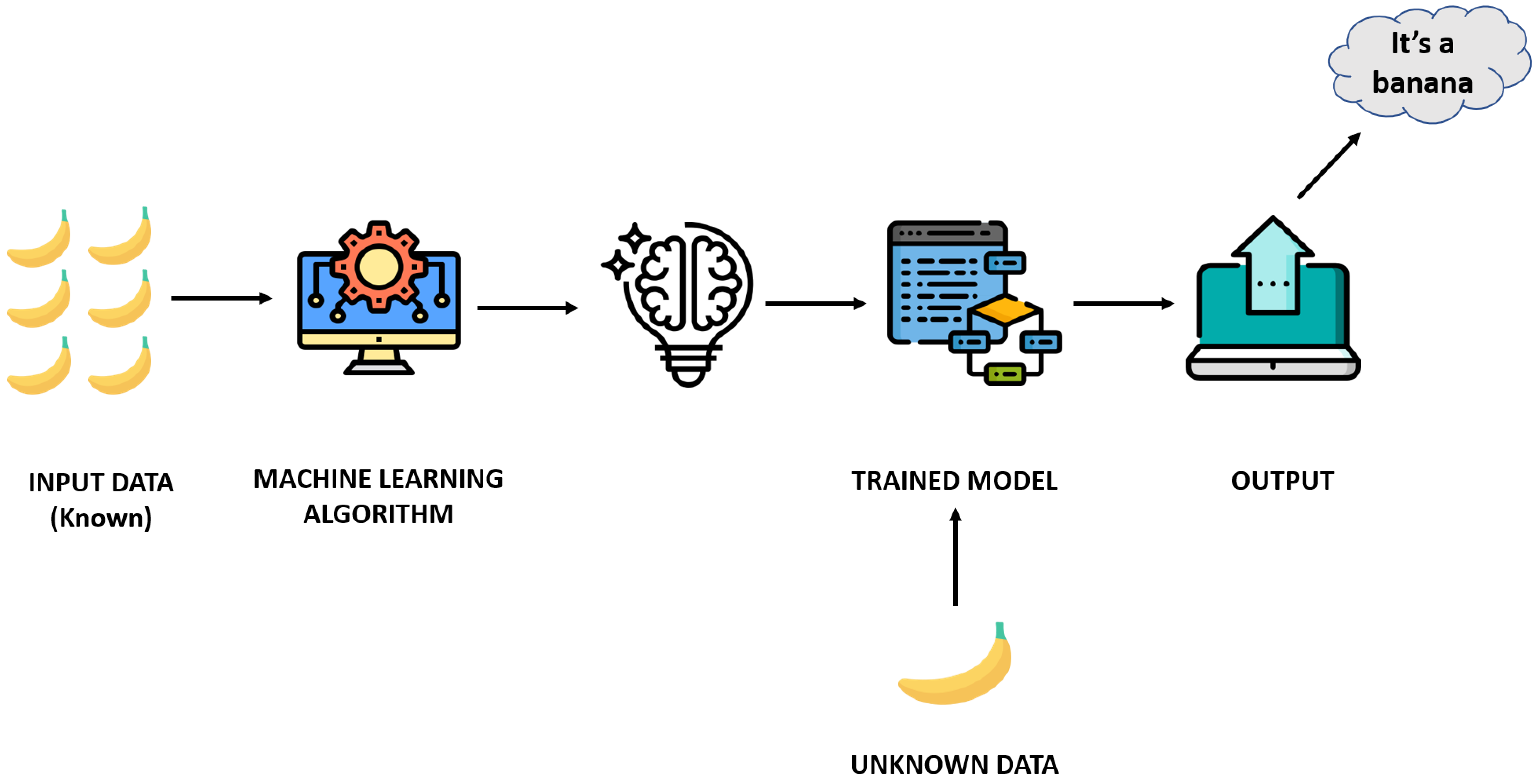
Supervised Machine Learning

Resources for Machine learning



Course slide from
DIKU 004 - Supervised Machine Learning.pdf

Supervised learning



How to work with Supervised machine learning

- Data collection
- Data processing
  - Data cleaning
  - Data labeling
  - Feature Engineering
  - Feature scaling
  - Data splitting (training, validation and test sets)
- Model Selection
- Training
- Validation
- Evaluation (with test data)
- Deploy the model
- Monitoring and mantainance

We will focus on models

Model types (supervised learning)

Classification

Support vector machines

Decision Trees

K-Nearest Neighbour

Random Forest

Logistic regression

Naïve Bayes classifier

…

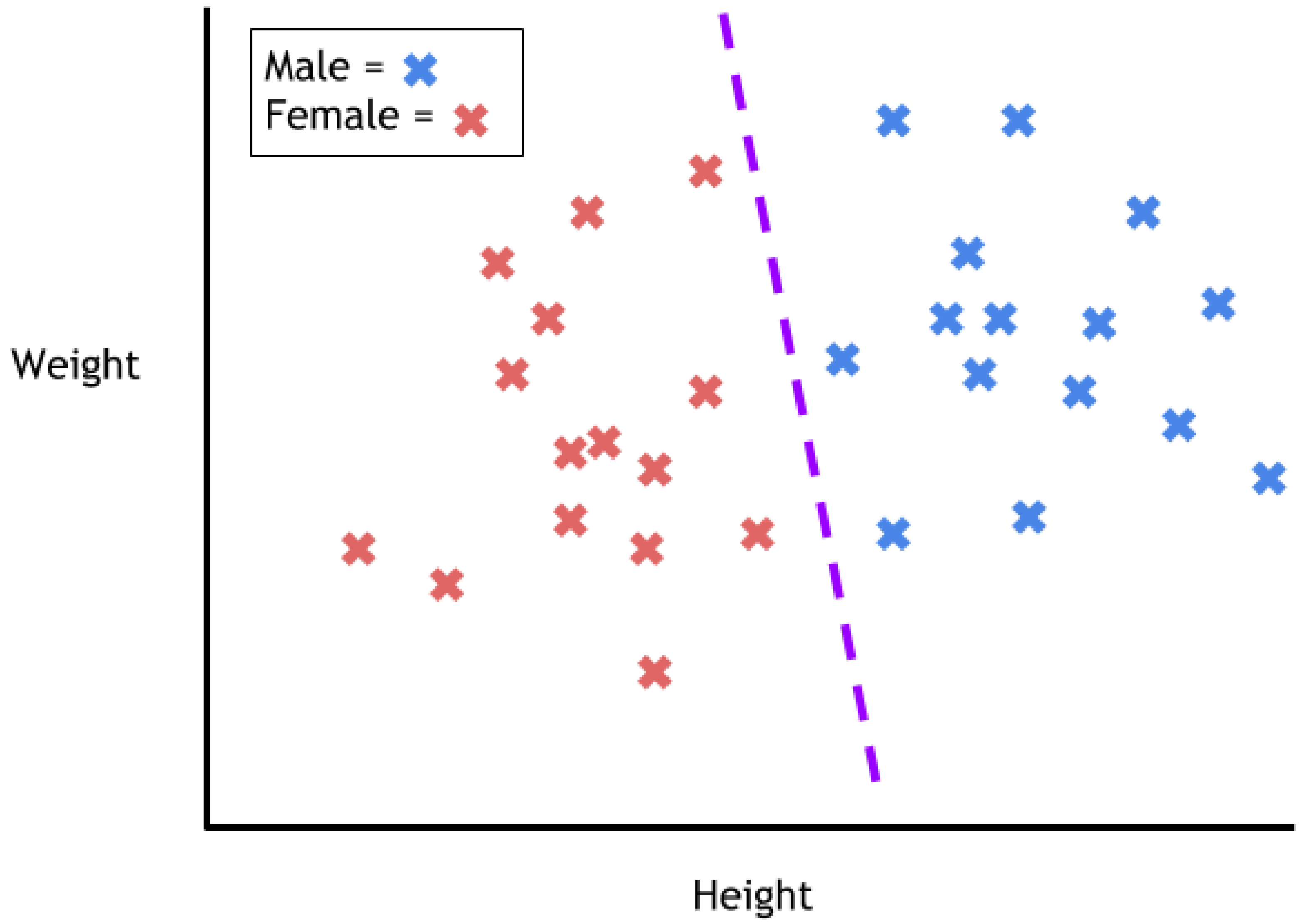
Regression

Linear Regression

Polynomial Regression

…

Classification

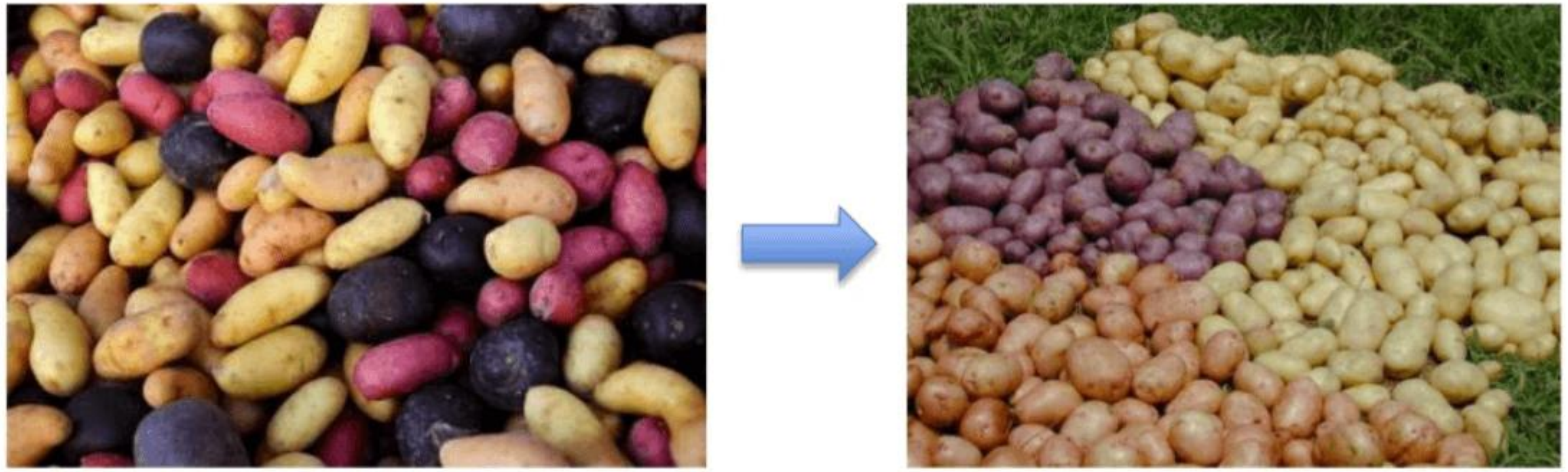


- Divides the data in classes / categories
- Use cases:
  - Spam detection (classification: spam or normal)
  - Analysis of the customer data to predict whether they will buy computer accessories (classification: Yes or No)
  - Classifying fruits from features like colour, taste, size, weight (classification: Apple, Orange, Cherry, Banana)
  - Gender classification from hair length (classification: Male or Female)
  - Stock market price prediction (classification: high or low)

Classification versus Clustering algorithms

Clustering

- Where we want to discover the groupings in data



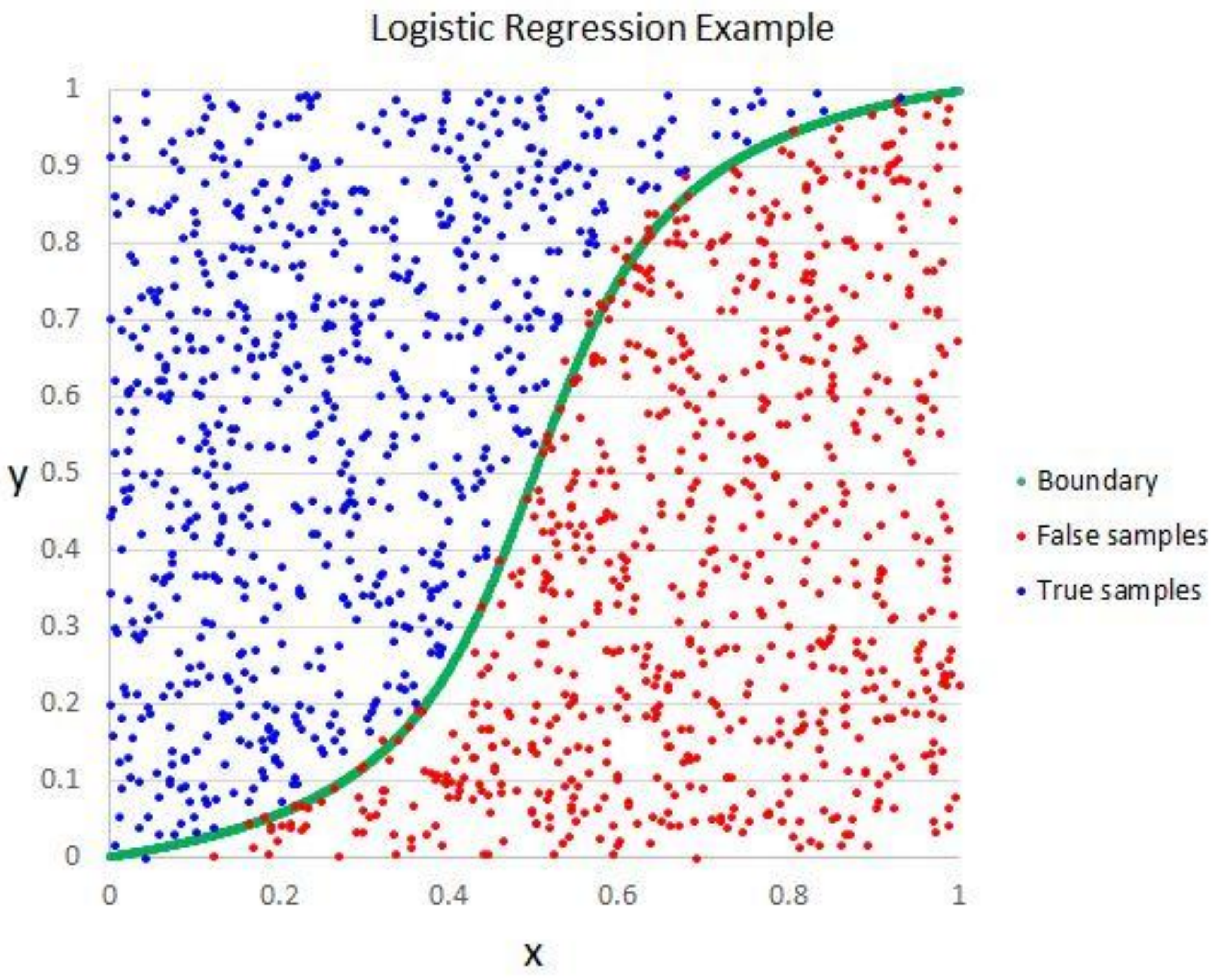
sample Cluster/group

Classification versus Clustering algorithms

- Classification is assigning labels
- Clustering is grouping data together
- Classification is supervised, clustering is unsupervised.

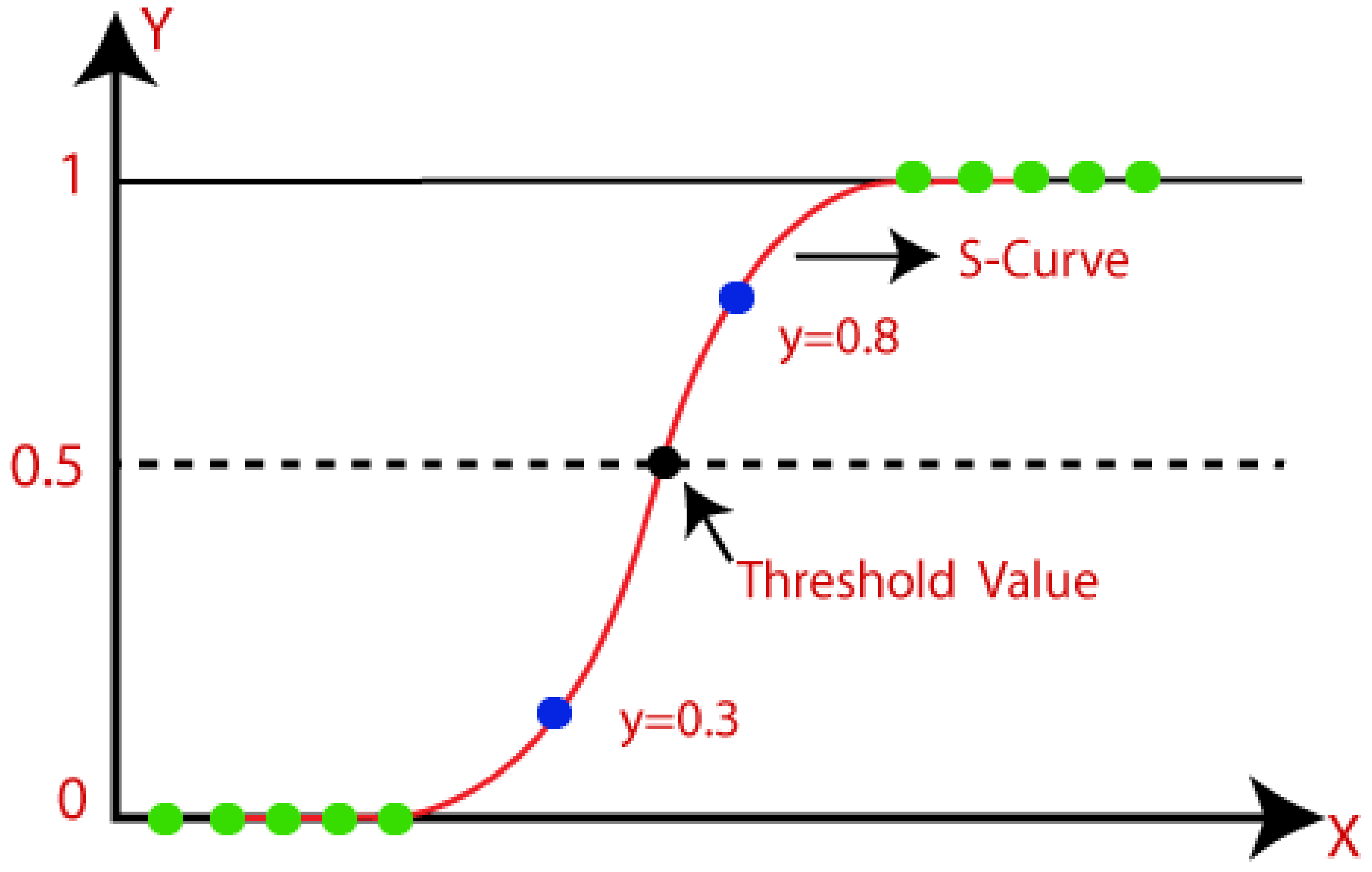
Classification -> Logistic regression

- is used when you have a classification problem - yes/no, pass/fail, win/lose, alive/dead, healthy/sick etc..
- It is the go-to method for binary classification problems
- The logistic function looks like a big S and will transform any value into the range 0 to 1.



Sigmoid function:

- Allows to put a threshold value. e.g 0.5 (use case spam detection)



Uses

- Logistic regression is used to predict the occurance of some event. e.g
  - Predict whether rain will occur or not
  - spam detection, Diabetes prediction, cancer detection etc.

Types of Logistic regression

- Binary logistic regression (e.g pass / fail)
- Multiclass logistic regression (e.g cats, dogs, sheep)
- Ordinal (low, medium, high)

Example run of an A.I algorithm

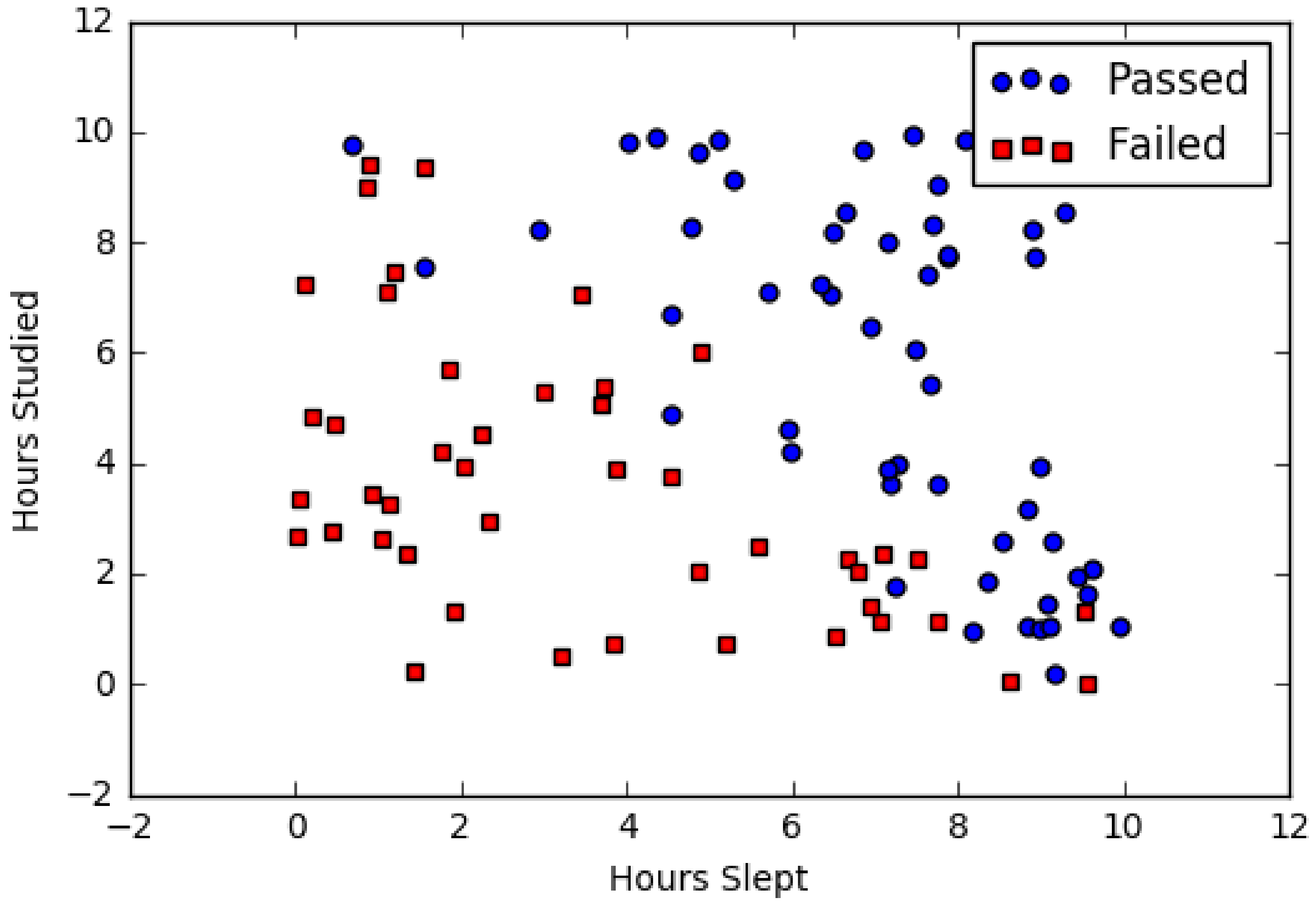
1. Develop the algorithm

2. Evaluate the algorithm

Develop (using binary logistic regression)

Step1 : Visualization

Studied	Slept	Passed
4.85	9.63	1
8.62	3.23	0
5.43	8.23	1
9.21	6.34	0

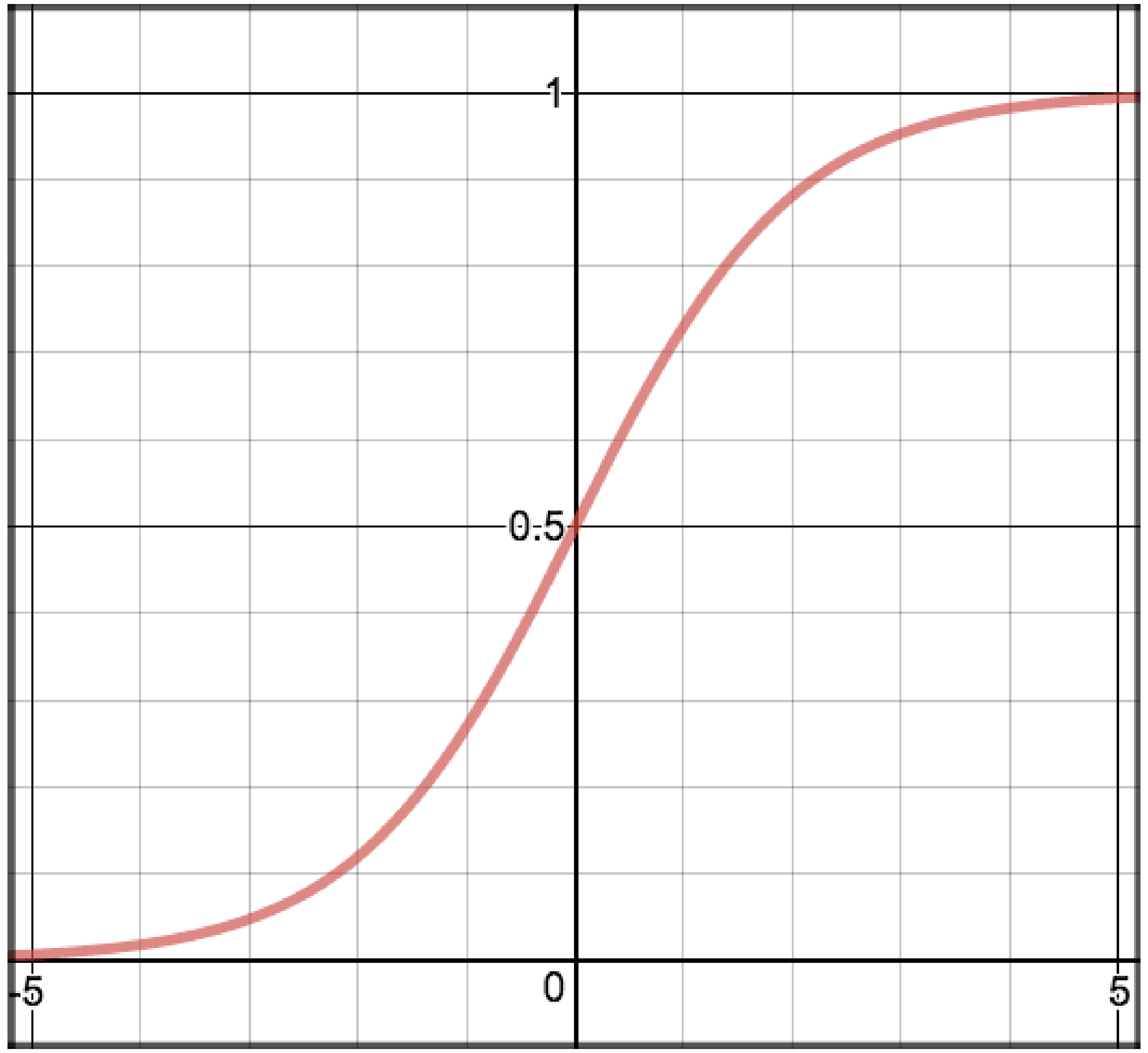


Step 2: Sigmoid activation (a bit of maths)

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

Code

```
def sigmoid(z):
  return 1.0 / (1 + np.exp(-z))
```

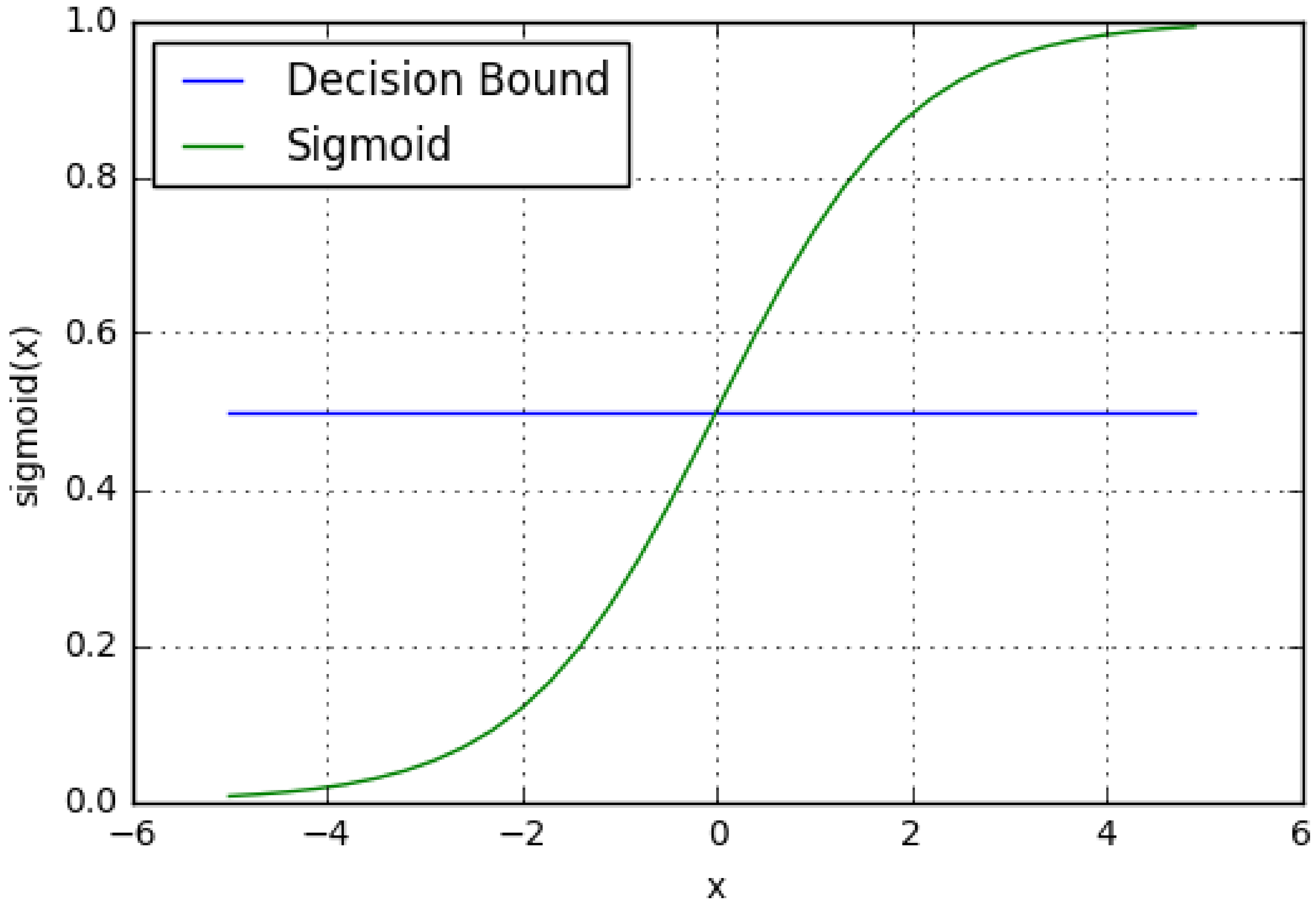


$s(z)$ = output between 0 and 1 (probability estimate)
z = input to the function (your algorithm's prediction e.g. mx + b)
e = base of natural log

Step 3: Decision boundary

- We need a threshold value

$$p \geq 0.5, class = 1$$
$$p < 0.5, class = 0$$



Step 4: Making predictions

- Using our knowledge of sigmoid functions and decision boundaries, we can now write a prediction function.
- A prediction function in logistic regression returns the probability of our observation being positive, True, or “Yes”.
- We call this class 1 and its notation is $P(class=1)$.

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

- If the model returns .4 it believes there is only a 40% chance of passing. If our decision boundary was .5, we would categorize this observation as “Fail.”“
- Code it

Second stage: evaluate the algorithm

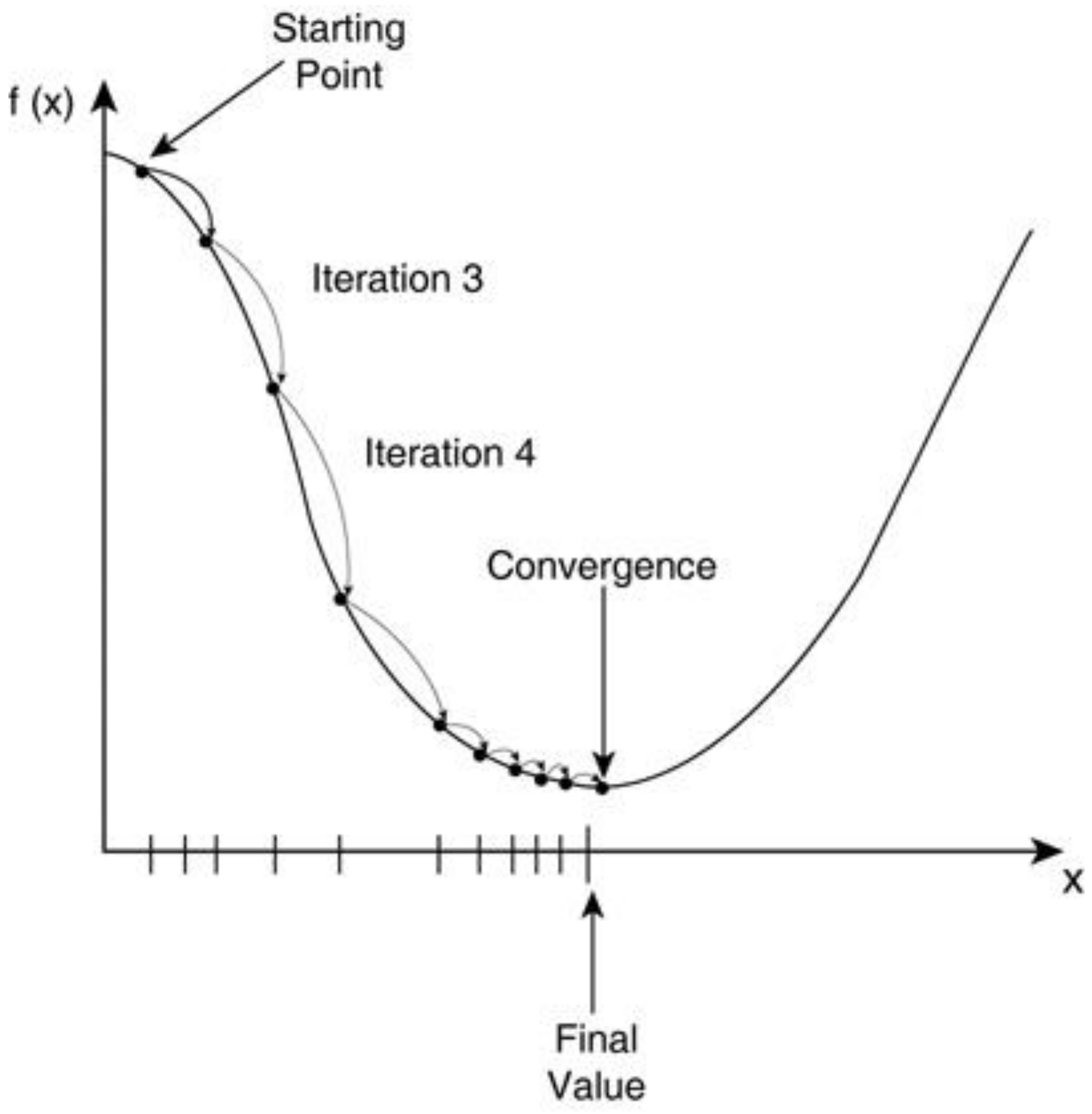
Evaluate using cost function

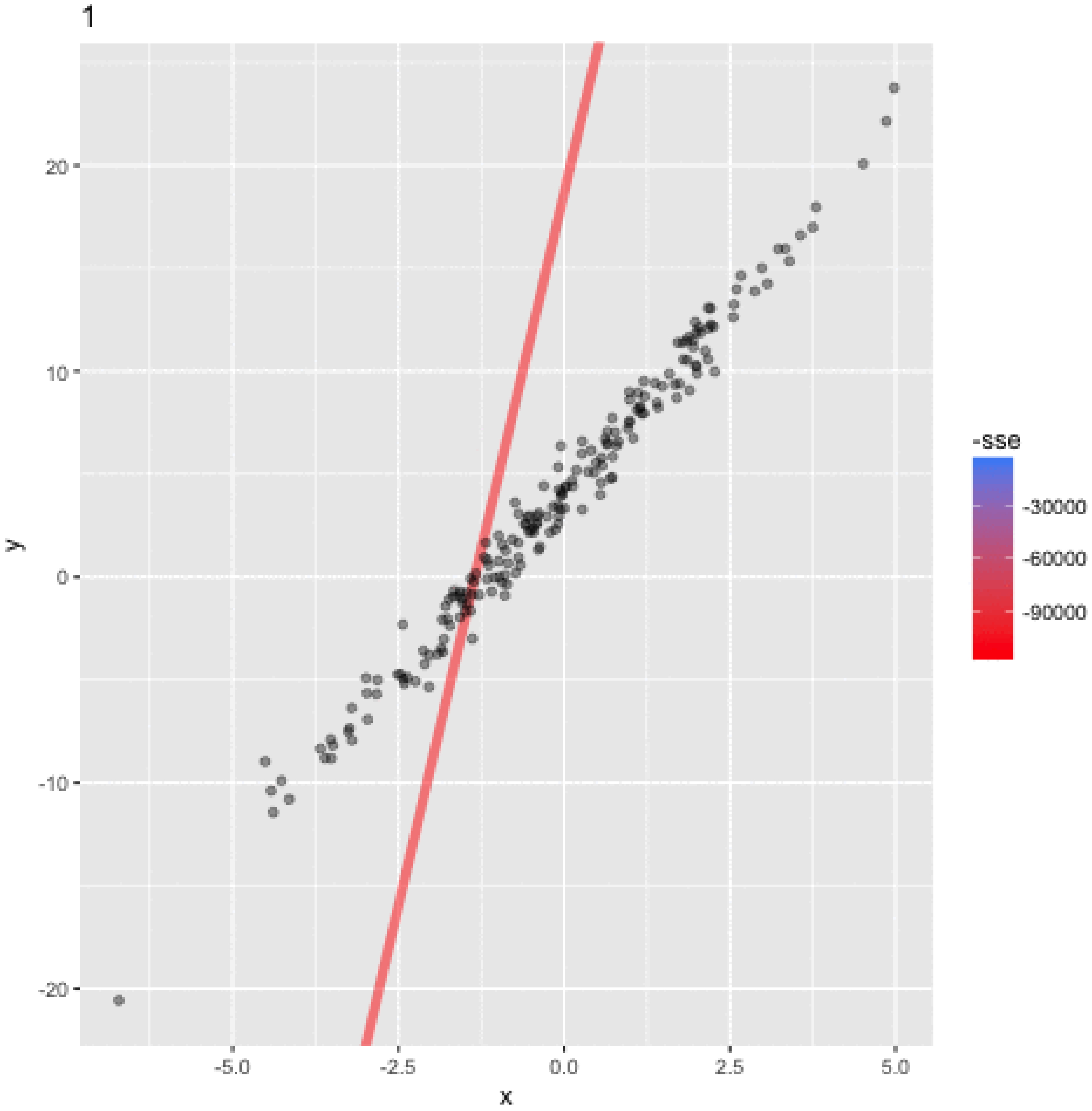
- What is a cost function ?
  - The cost function helps the learner to correct / change behaviour to minimize mistakes.
- In ML, cost functions are used to estimate how badly models are performing.
- Put simply, *a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and y.*

Cost function: Gradient descent

- Gradient Descent is an optimization algorithm used to tweak parameters iteratively to minimize the cost function.

- Gradient descent enables a model to learn the gradient or *direction* that the model should take in order to reduce errors (differences between actual y and predicted y).




1
20
10
0
-10
-20
y
-5.0
-2.5
0.0
2.5
5.0
x
-sse
-30000
-60000
-90000

Develop: Map probabbilities to classes

- The final step is assign class labels (0 or 1) to our predicted probabilities.

Decision boundary

```
def decision_boundary(prob):
  return 1 if prob >= .5 else 0
```

Convert probabilities to classes

```
def classify(predictions):
  '''
  input  - N element array of predictions between 0 and 1
  output - N element array of 0s (False) and 1s (True)
  '''
  decision_boundary = np.vectorize(decision_boundary)
  return decision_boundary(predictions).flatten()
```

Example output

```
Probabilities = [ 0.967, 0.448, 0.015, 0.780, 0.978, 0.004]
Classifications = [1, 0, 0, 1, 1, 0]
```

Final stage: Finish it up

- Train the model
- Evaluate the model
  - Minimzie the cost with repeated iterations
- Measure the accuracy of your outputs
- Meausre the probabbility score

Classification -> Naive Bayes classification algorithm

- A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- e.g a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter.
- Its an easy to build algorithm and a very powerful one.

BAYES THEOREM

Bayes's theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event

Probability of B occurring given evidence A has already occurred

Probability of A occurring

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|A) \cdot P(A)}{P(B)}$$

Probability of A occurring given evidence B has already occurred

Probability of B occurring