Designing a Machine Learning Model

EXPLORING APPROACHES TO MACHINE LEARNING

Janani Ravi
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

Data-driven decisions and actions

Rule-based approaches to learning

Learning dynamically from changing data using machine learning

Feature extraction from unstructured data using deep learning

Traditional machine learning vs. deep learning

Prerequisites and Course Outline

Prerequisites

Comfortable programming in Python

Some familiarity with ML models

Understanding of basic math

- Mean, standard deviation

Prerequisite Courses

Understanding Machine Learning with Python

How to Think About Machine Learning Algorithms

Building Your First scikit-learn Solution

Course Outline

Approaches to machine learning

Choosing the right machine learning problem

Choosing the right machine learning solution

Building simple ML solutions

Designing machine learning workflows

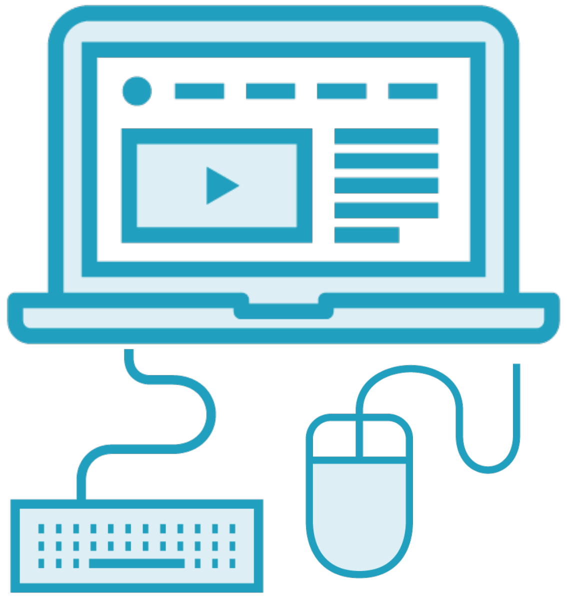
Building ensemble learning and neural network solutions

Case Study: Sentiment Analysis

Changing Patterns of Online Behavior

“Surf/Browse”

c. 1990 - c. 2000



“Search-Find-Obtain”

c. 2000 - c. 2008

“Share-Discover”

c. 2008 - Present

Share-Discover

Always online

Share with network

Discover through network

Stream of online opinions

Opinions Contain Information

Reviews

Tweets and Posts

Messages

Swipes

Data Analyst

Collect opinions

Extract information from them

Act on that information

Changing Patterns of Online Behavior

Collect Opinions

Scrape/harvest comments, articles, tweets...

Extract Information

Perform sentiment analysis

Act

Buy/sell stocks, target advertising spend,...

Collect Opinions

Researchers use public datasets

Companies use proprietary data

Scrapers use media signals

“Big Data”

Unstructured data

Extract Information

Tag data item with values for sentiments

One/more categorical data series created

Analyze categorical data

Extract Information

Data item to analyze

Tweet, email, message, review, …

Sentiment identified

“Positive”, “Negative”,”Neutral”

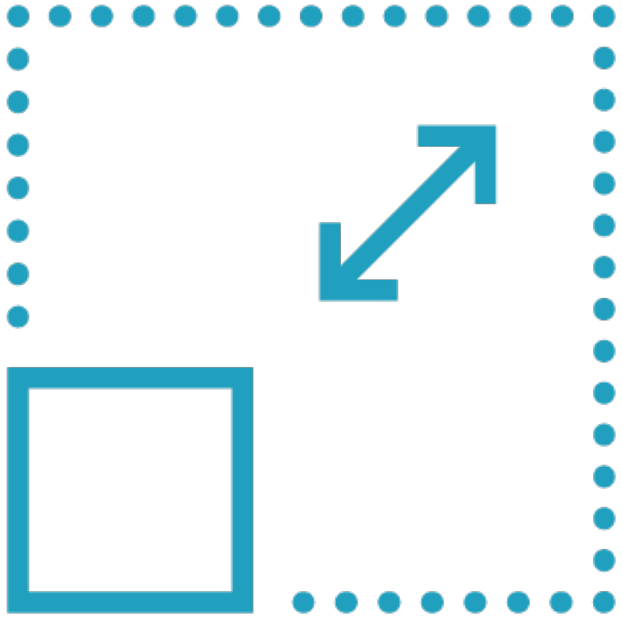
Categorical variable

+1, 0, -1

Analyzing Categorical Sentiment Data

Logistic Regression

Relationships between variables



Quadrant Analysis

Clusters of data with similar characteristics

Act

Trade financial markets

Change or reallocate ad budgets

Tailor electoral strategy

Decide product recall strategies

Changing Patterns of Online Behavior

Collect Opinions

Scrape/harvest comments, articles, tweets...

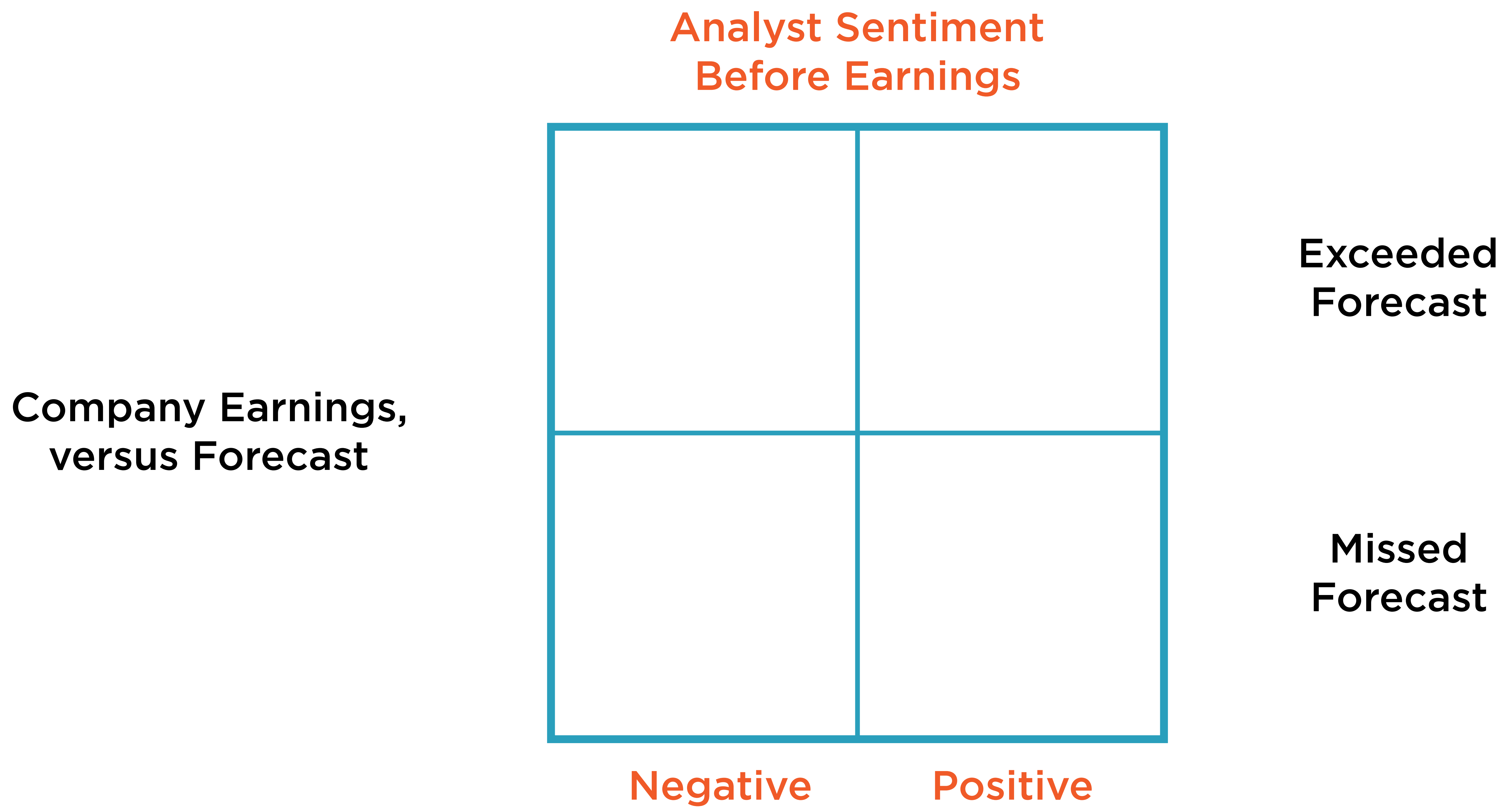
Extract Information

Perform sentiment analysis

Act

Buy/sell stocks, target advertising spend,...

Sentiment Analysis in Event-Driven Trading



Sentiment Analysis in Event-Driven Trading

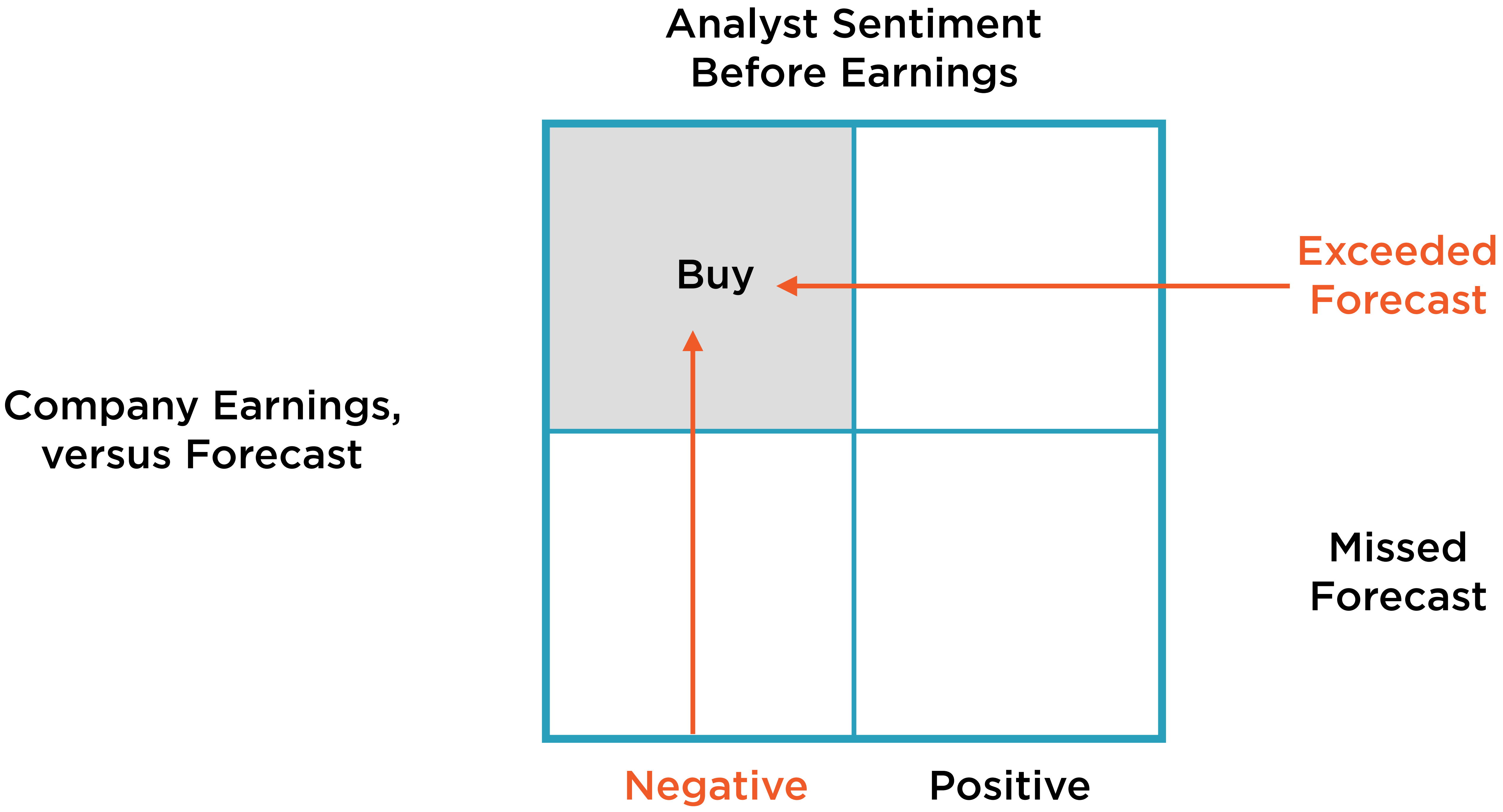
Company Earnings Releases

Better or worse than analyst expectations?

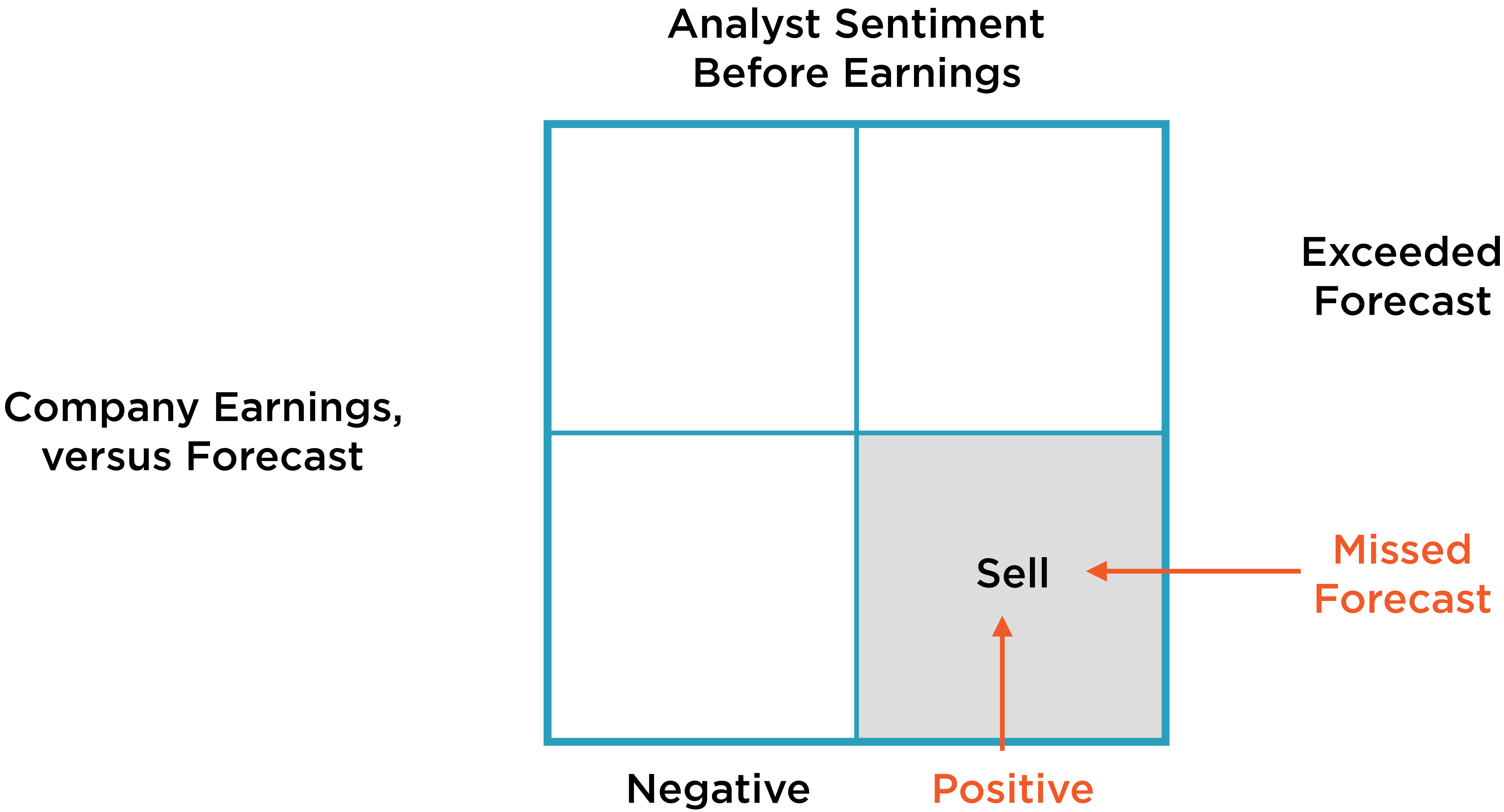
Financial Traders

Buy or sell?

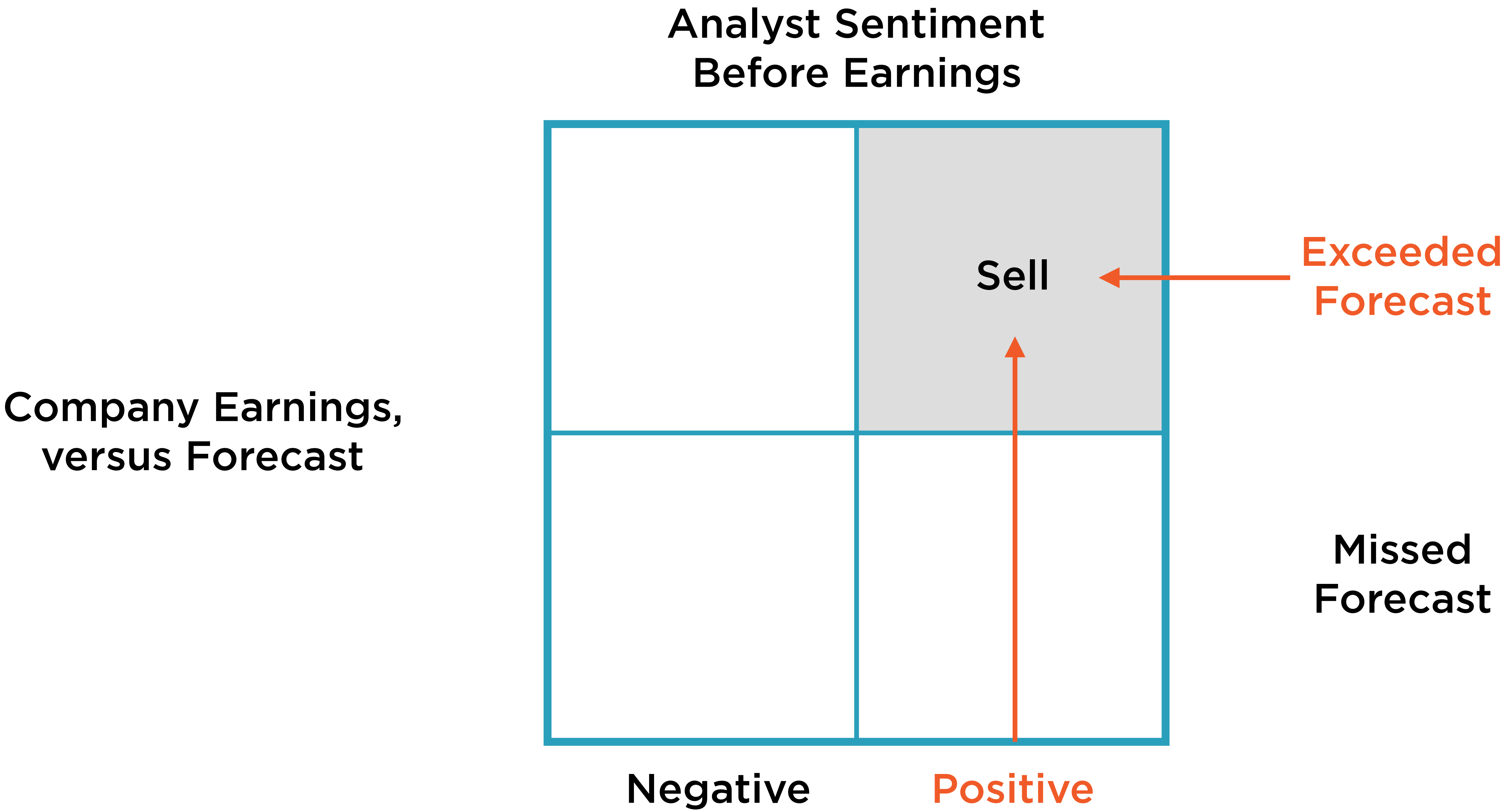
Sentiment Analysis in Event-Driven Trading



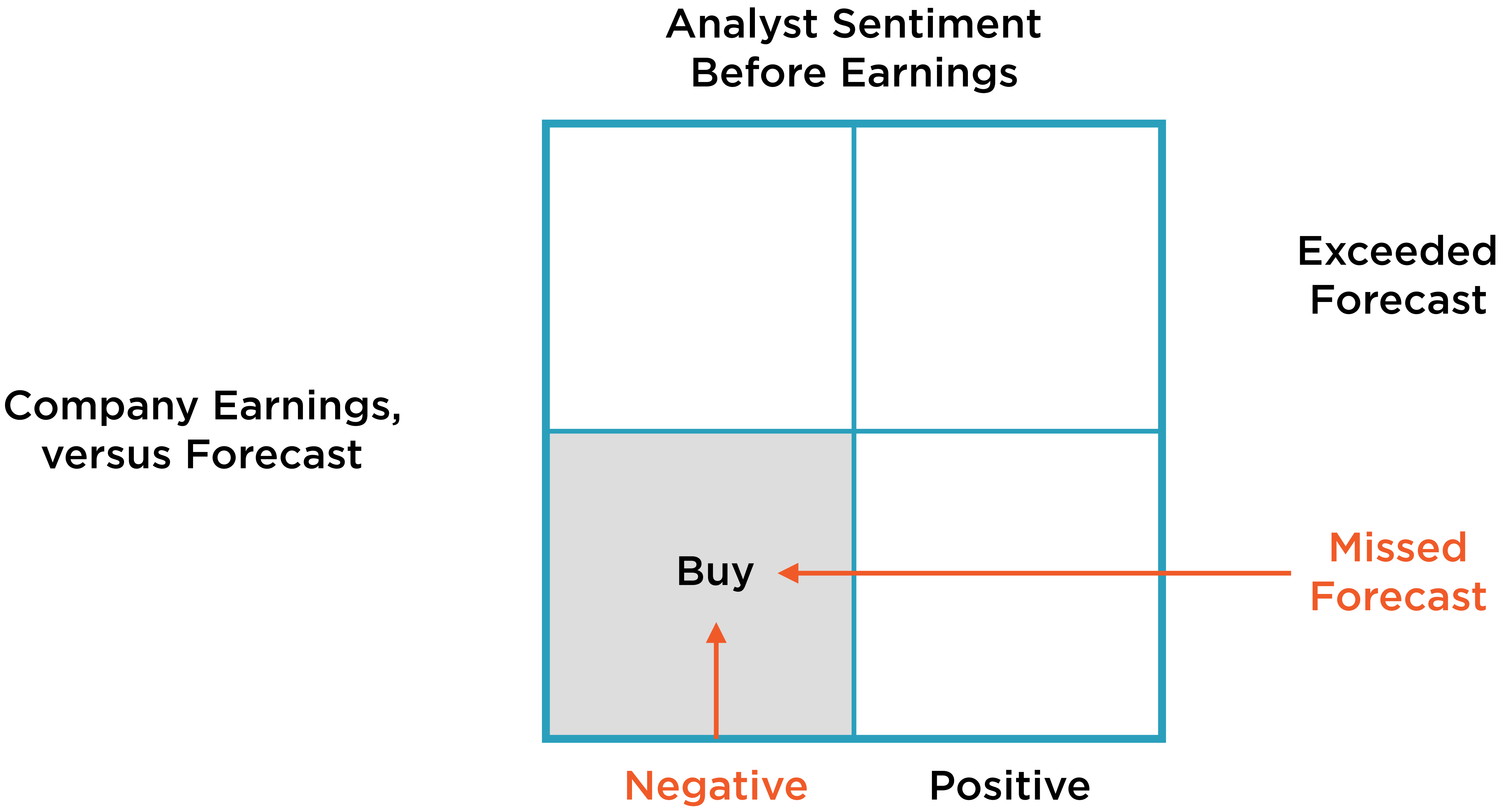
Sentiment Analysis in Event-Driven Trading



Sentiment Analysis in Event-Driven Trading



Sentiment Analysis in Event-Driven Trading



Insight: “Buy the Rumor, Sell the News”

Buy the rumor

If market sentiment was negative, buy even if earnings are poor

Sell the news

If market sentiment was positive, sell even if earnings are great

Polarity Detection for Sentiment Analysis

Sentiment Analysis Systems

Polarity

Positive or negative?

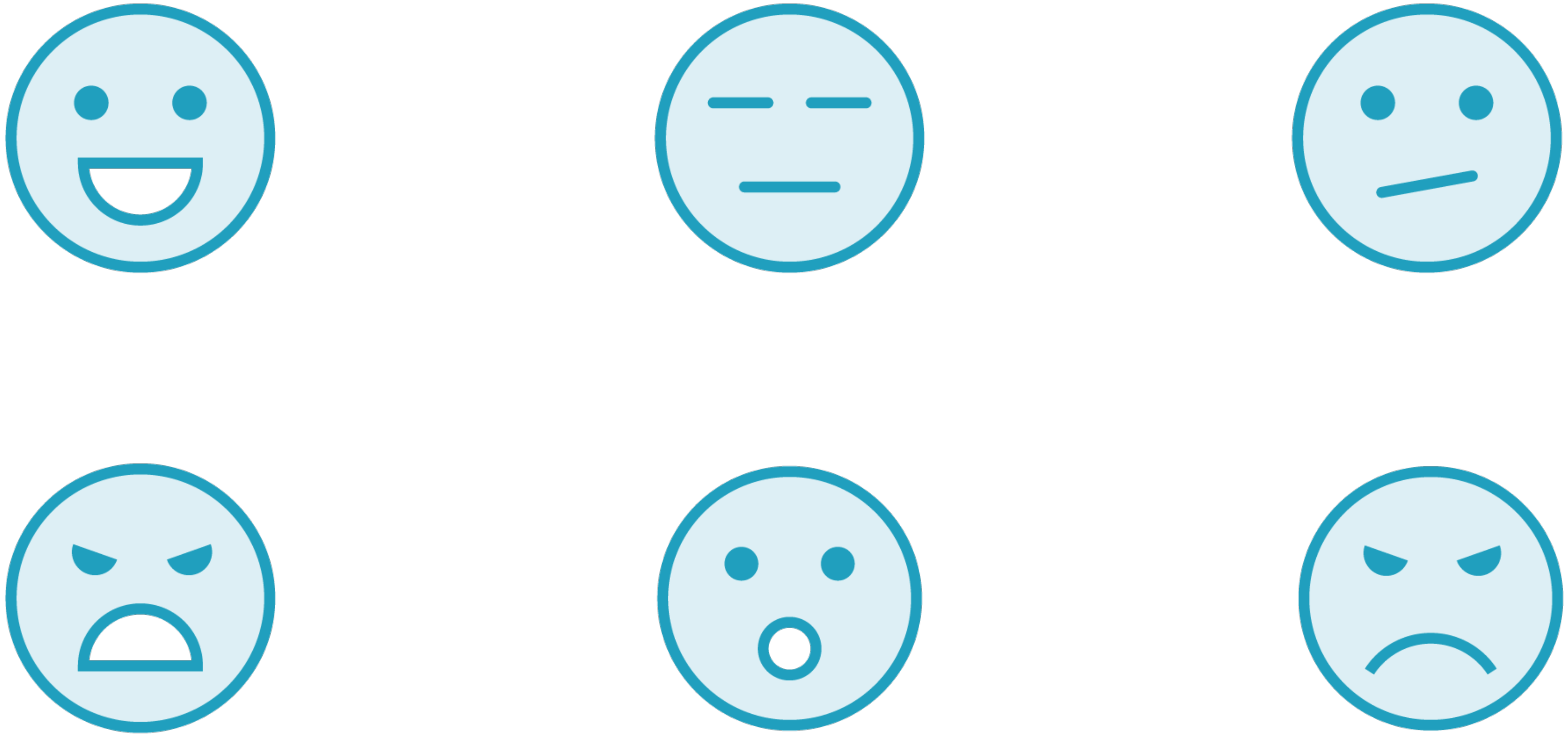
Subjectivity

Subjective or objective?

Aspects

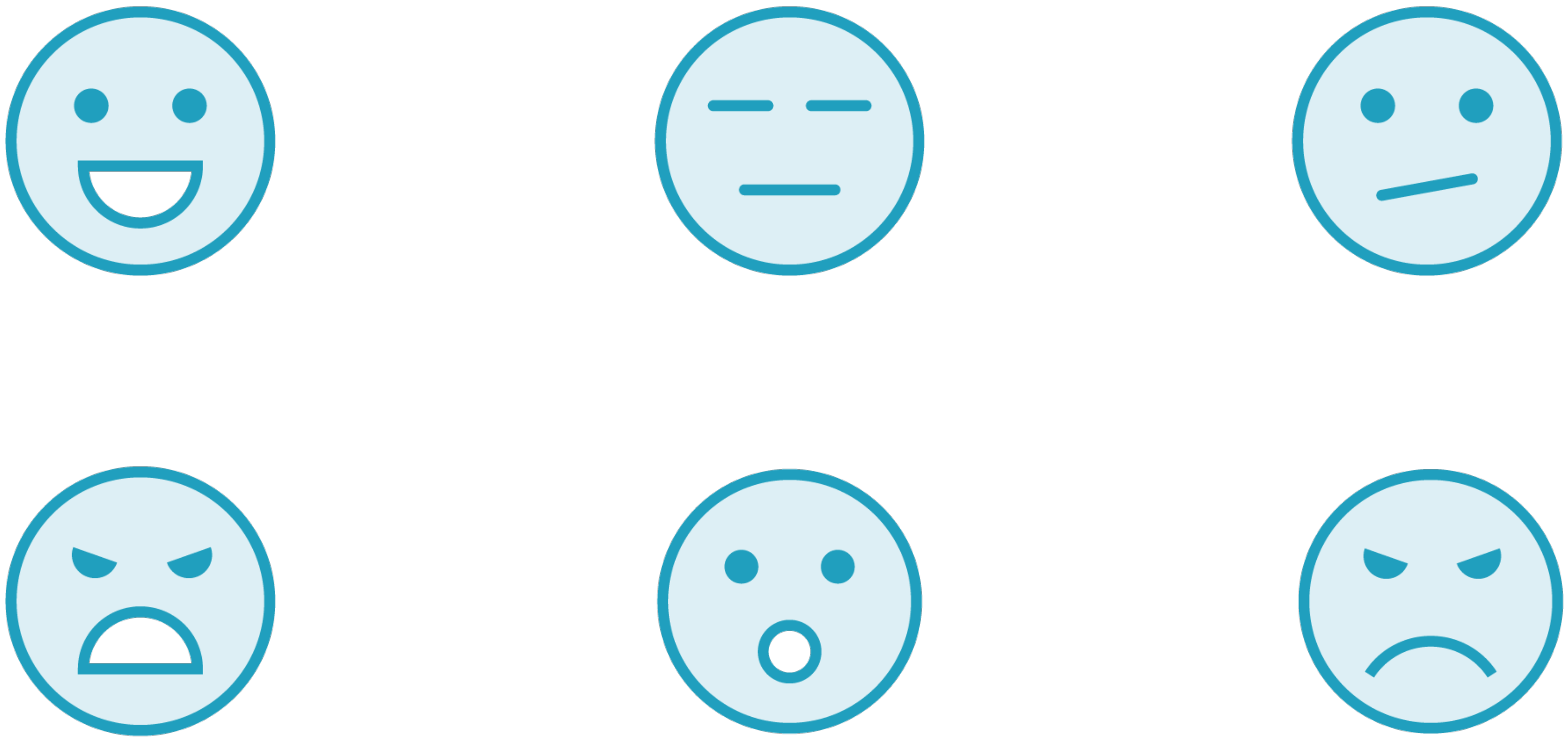
Part or whole?

Opinions Are Very Complex



But sentiment analysis need not be (if we set up the problem right)

Either-or Decisions Are Simple



Human brains are very efficient at making binary decisions

Binary Decisions

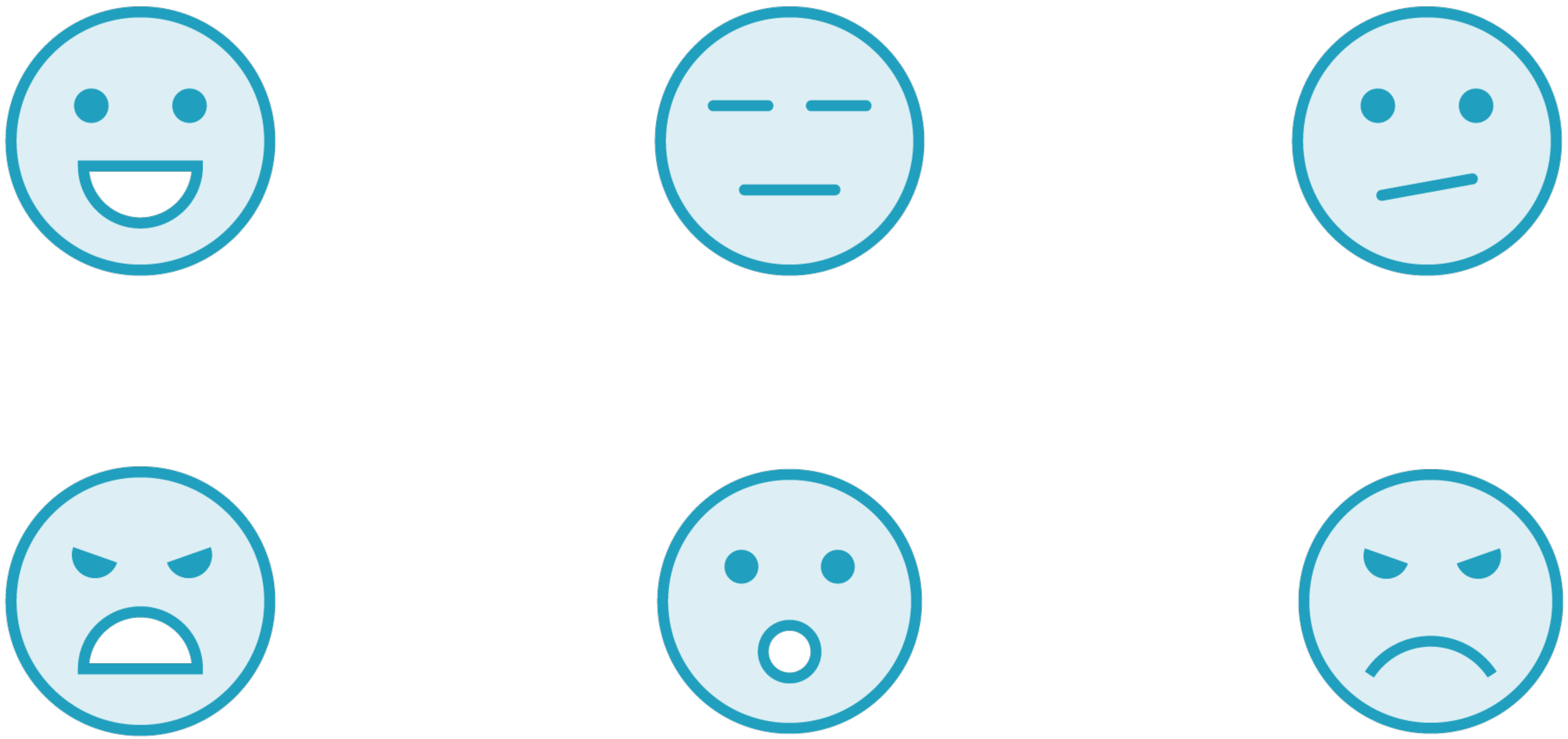
Hot or not?

Buy or sell?

Fight or flight?

For or against?

Opinions Are Very Complex



Model sentiment analysis as a
Binary Classification problem

Binary Classification

Positive

Not Positive

Model sentiment analysis as a Binary Classification problem

Binary Classification

Positive

Not Positive

Binary classification is a well-studied, well-understood problem

Binary Decisions

Binary Decisions

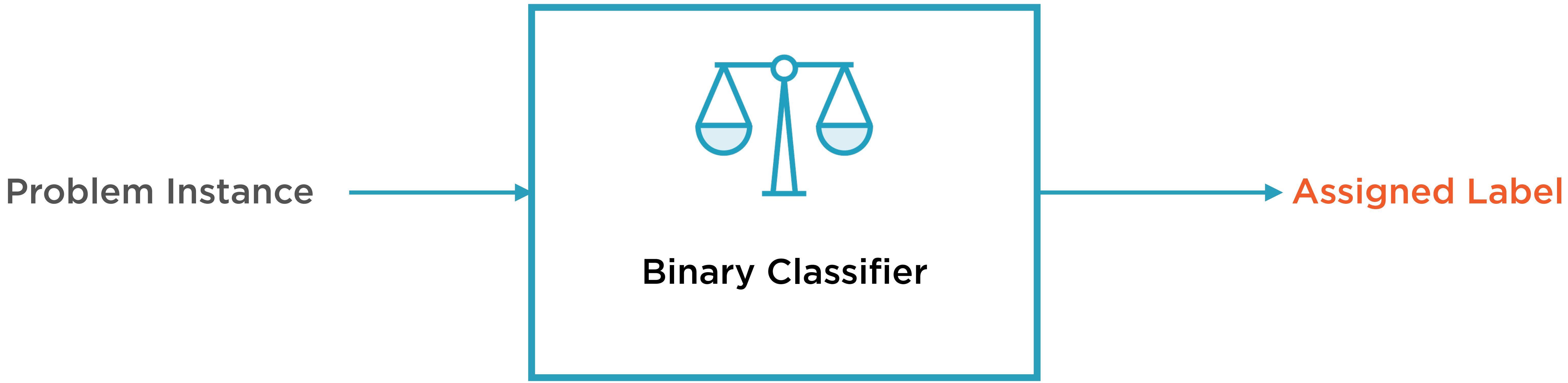
Comment: Positive or negative?

Email: Spam or ham?

Transactions: Fraud or legit?

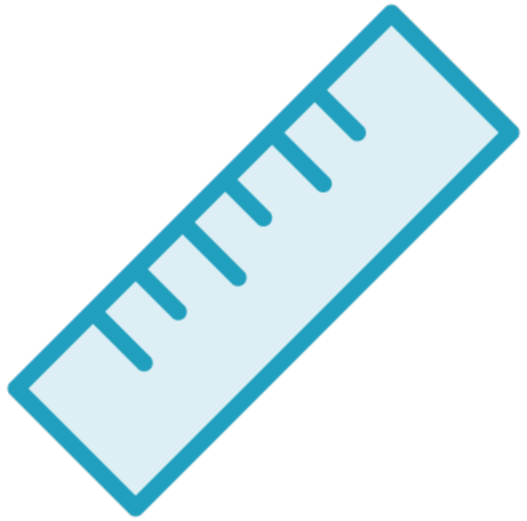
Rule-based and ML-based Binary Classifiers

Sentiment Analysis as Binary Classification



The binary classifier is a function that takes in a problem instance, and assigns a label

Binary Classifiers



Rule-based Classifiers

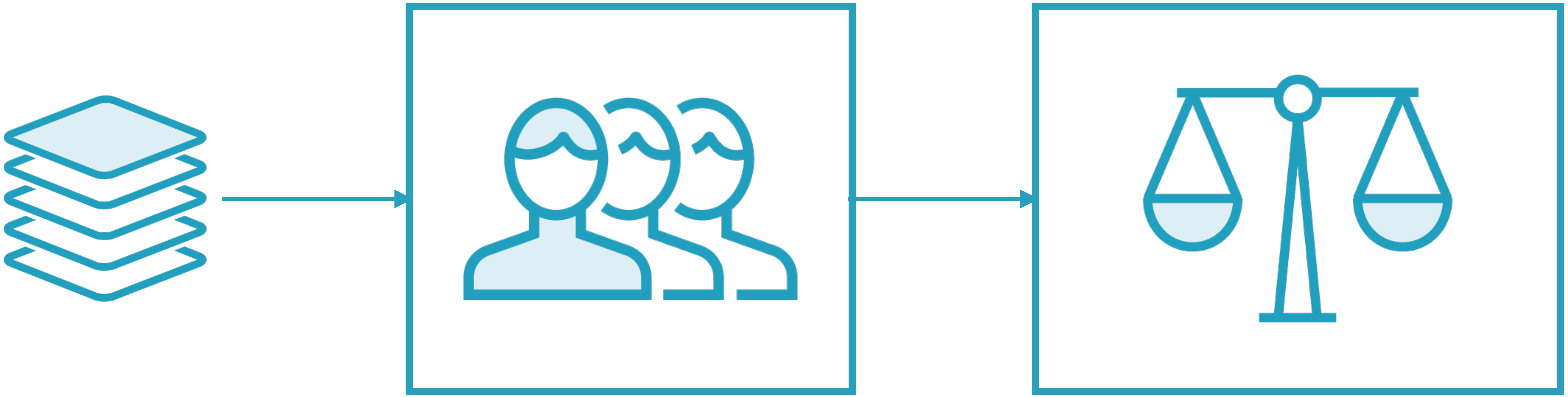
Rules drawn up by experts are used to assign a label to problem instance

ML-based Classifiers

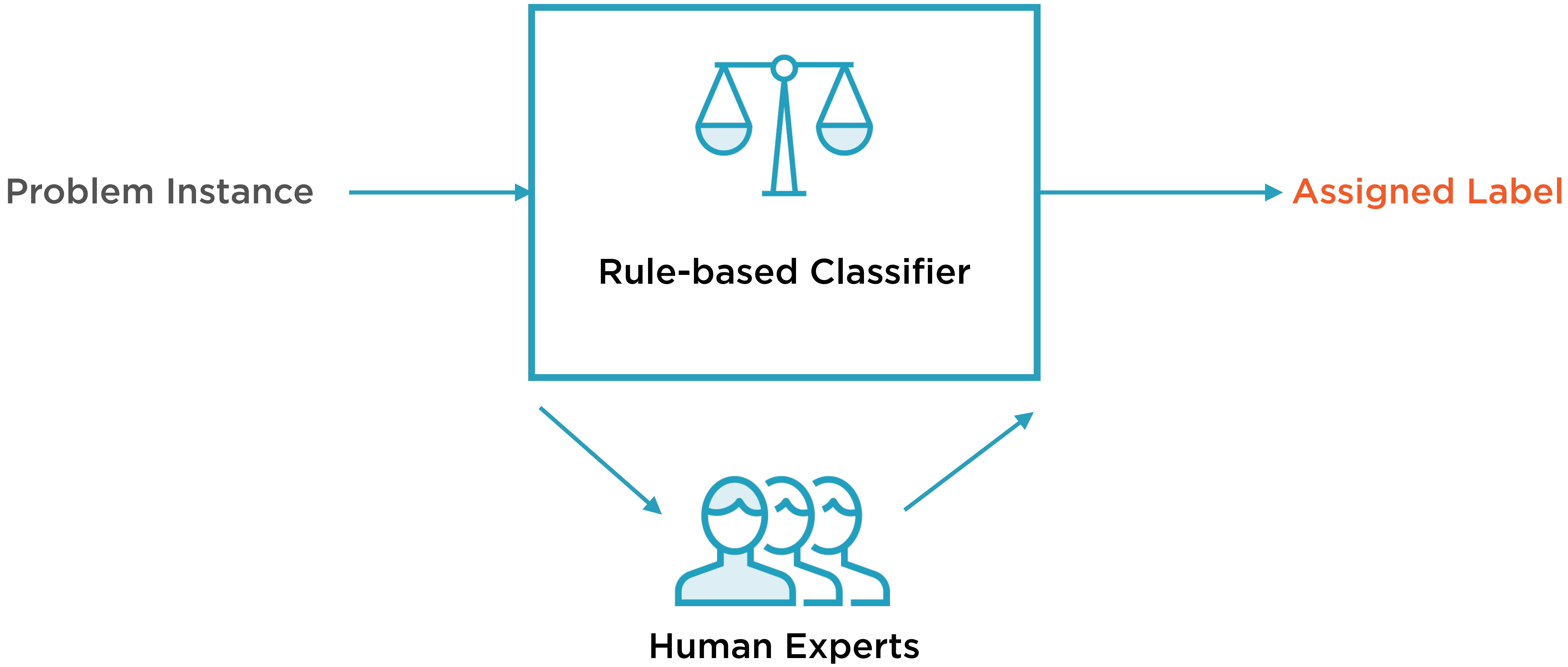
Label is assigned based on patterns displayed in aggregate data

Rule-based Binary Classifier

Rule-based Binary Classifier

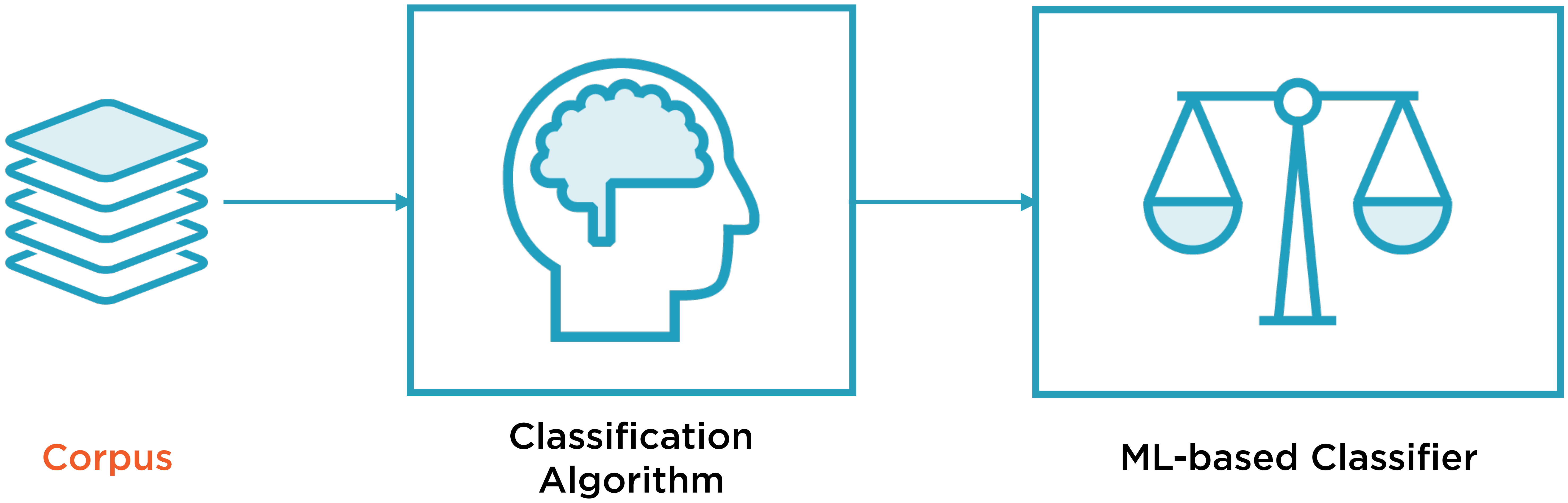


Rule-based Binary Classifier

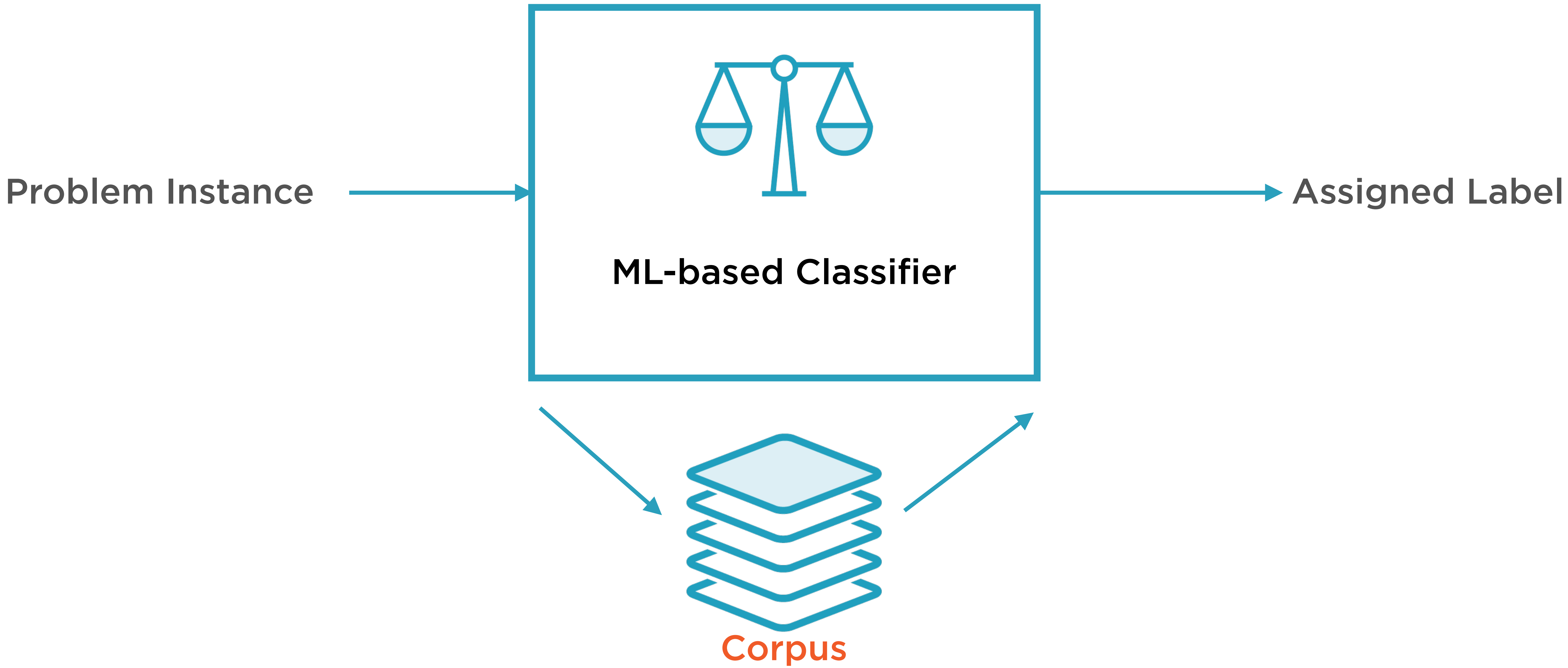


ML-based Binary Classifier

ML-based Binary Classifier



ML-based Binary Classifier



ML-based and Rule-based Classifiers

ML-based

Dynamic - alter output based on patterns in data

Far less need for expert skill

To update classifier, simply update corpus

Rule-based

Static - rules are applied independent of data being analyzed

Experts vital for formulating rules

To update classifier, need to update rules i.e. recode model

ML-based and Rule-based Classifiers

ML-based	Rule-based
Large, high-quality data corpus crucial	No corpus required
Cannot operate in isolation on a single problem instance	Can operate on isolated problem instances
Explicit training step	No training step required

Rule-based Analysis

- Problem statement is fairly simple
- Rules are straightforward and can be easily codified
- Rules change infrequently
- Few problem instances to train ML models

ML-based Analysis

Problem statement is reasonably complex

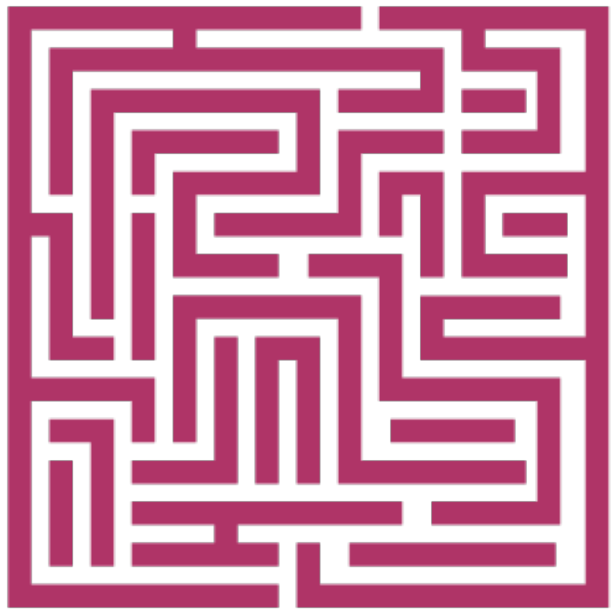
Hard to find patterns using visualizations and other exploratory tools

Decision variables sensitive to data, need to change as new information is received

Large corpus available to train models

Understanding Machine Learning

Machine Learning



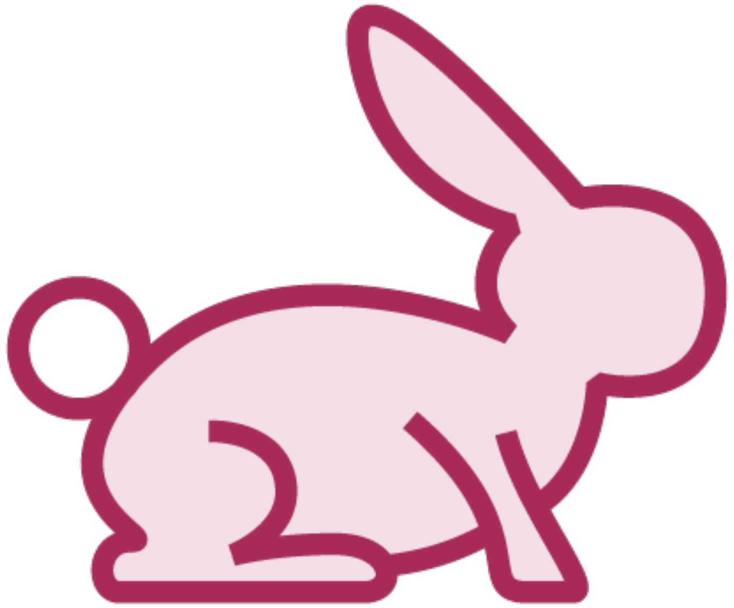
Work with a huge maze of data

Find patterns

Make intelligent decisions

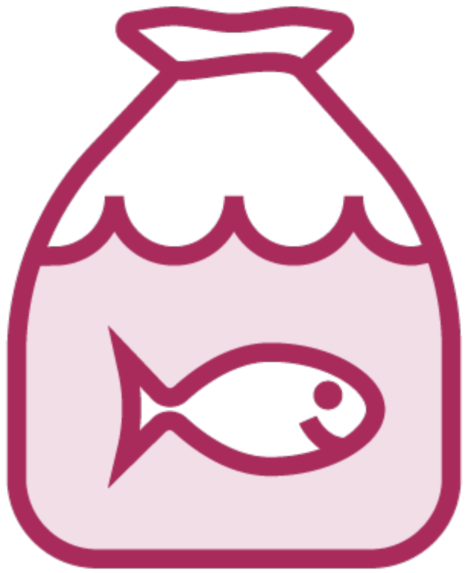
A machine learning algorithm is an algorithm that is able to learn from data

Whales: Fish or Mammals?



Mammals

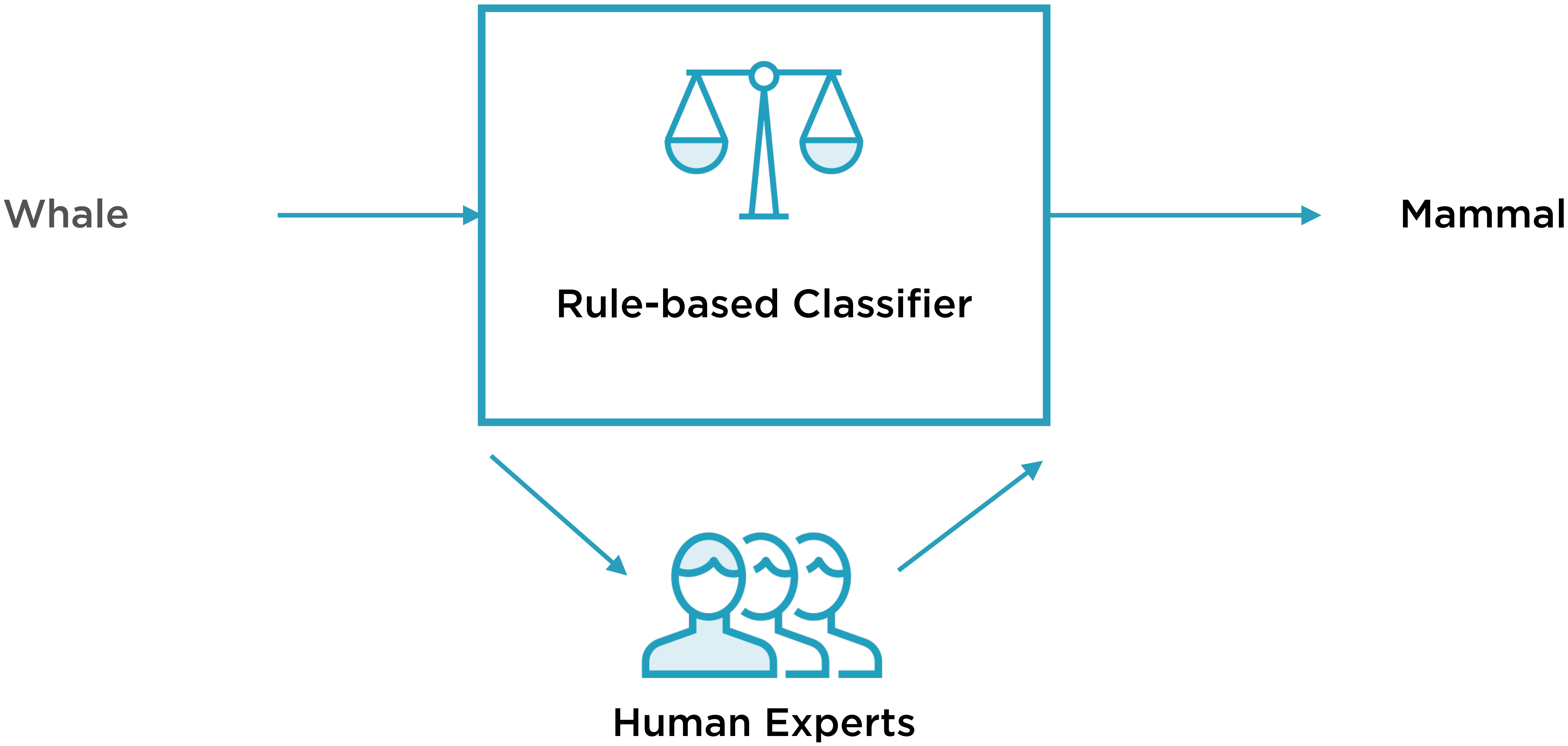
Members of the infraorder *Cetacea*



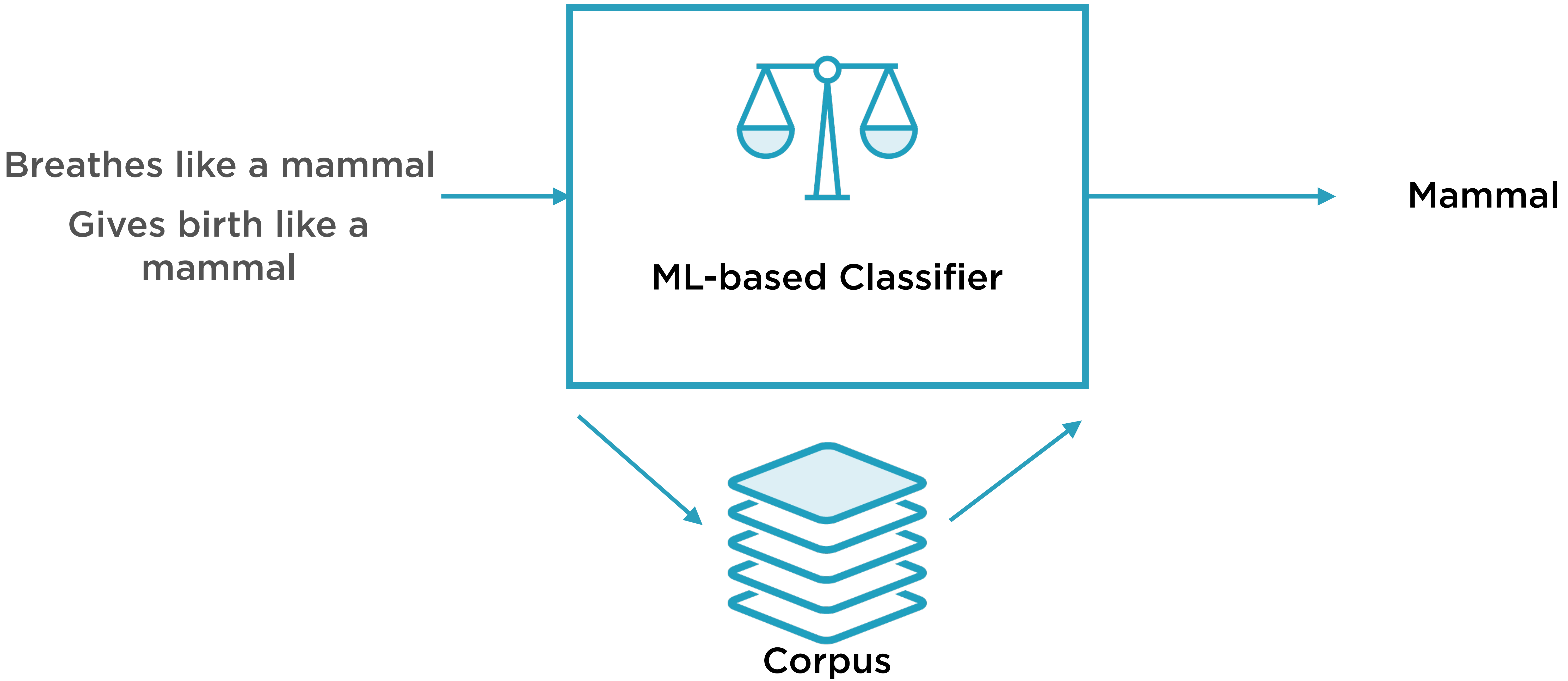
Fish

Look like fish, swim like fish, move with fish

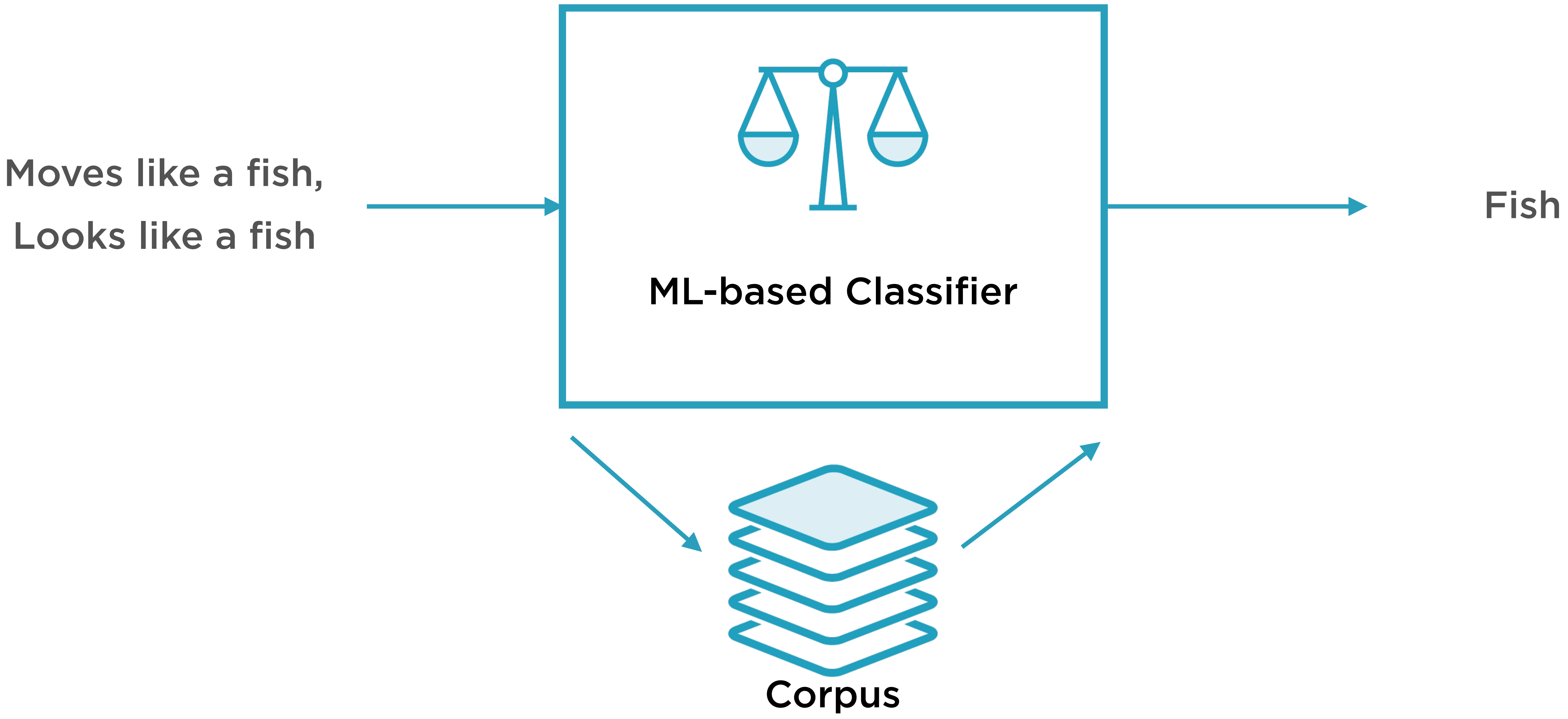
Rule-based Binary Classifier



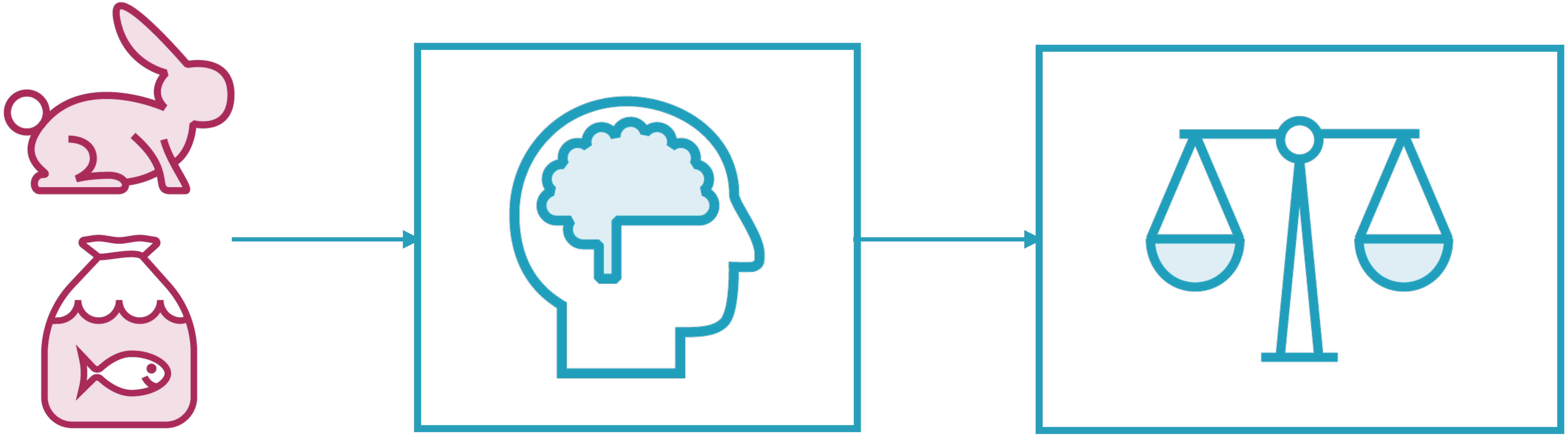
ML-based Binary Classifier



“Traditional” ML-based Binary Classifier



ML-based Binary Classifier



Corpus

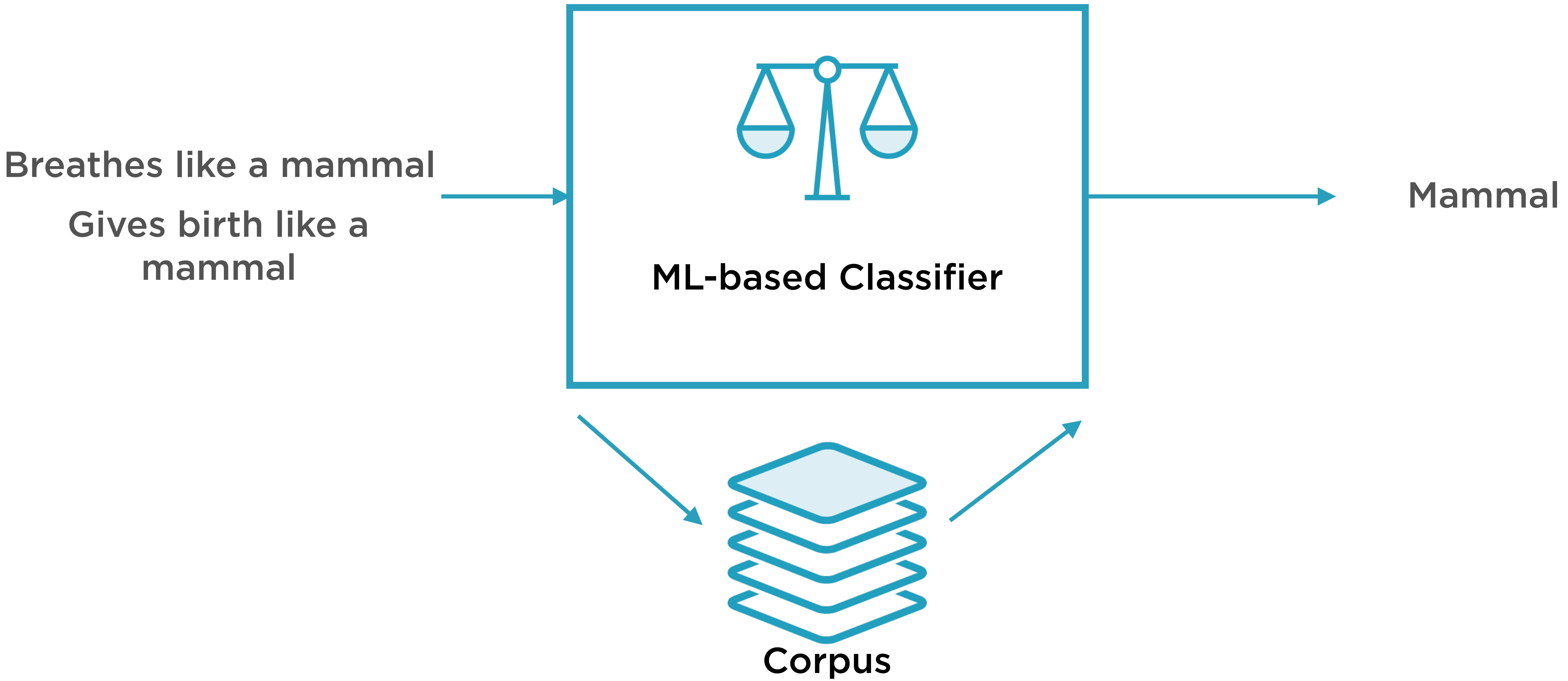
Classification Algorithm

ML-based Classifier

ML-based Binary Classifier

ML-based	Rule-based
Dynamic	Static
Experts optional	Experts required
Corpus required	Corpus optional
Training step	No training step

“Traditional” ML-based Binary Classifier



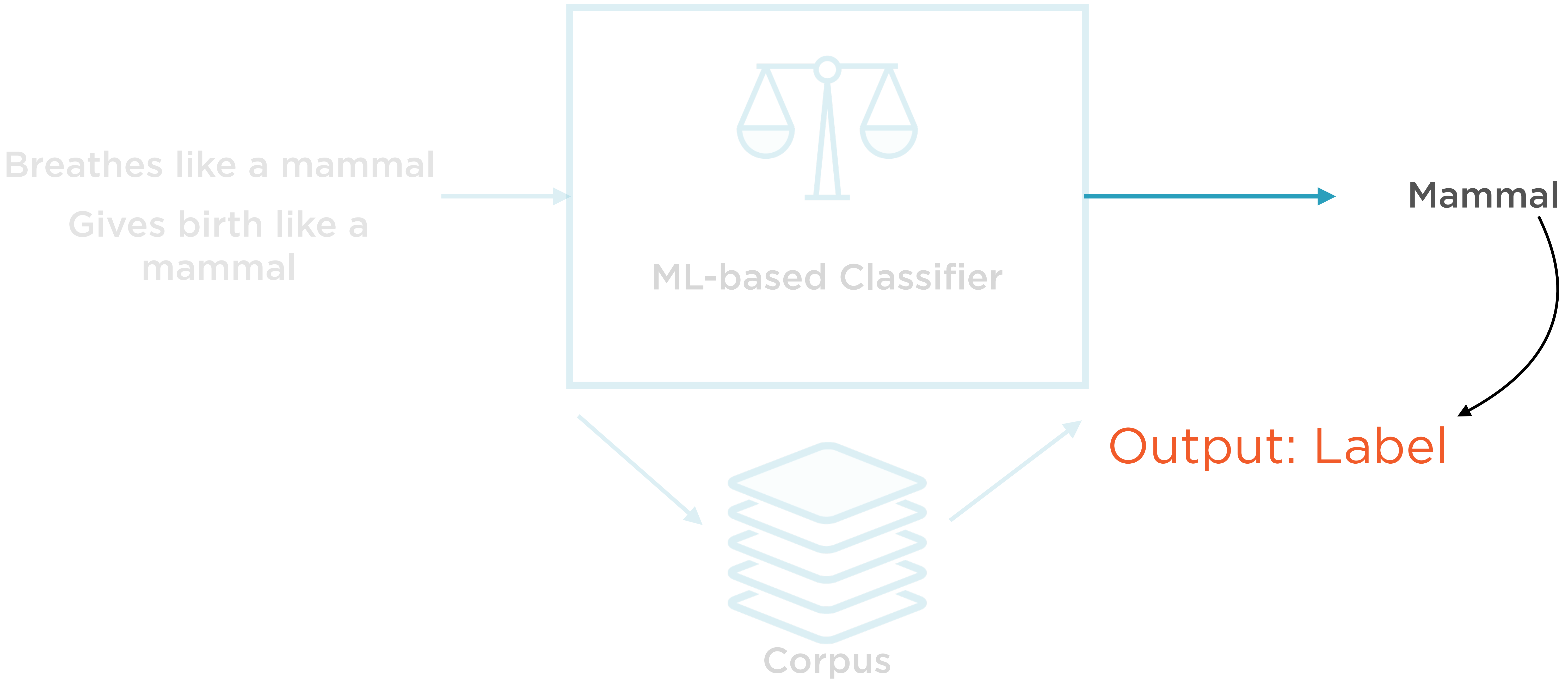
“Traditional” ML-based Binary Classifier

Breathes like a mammal

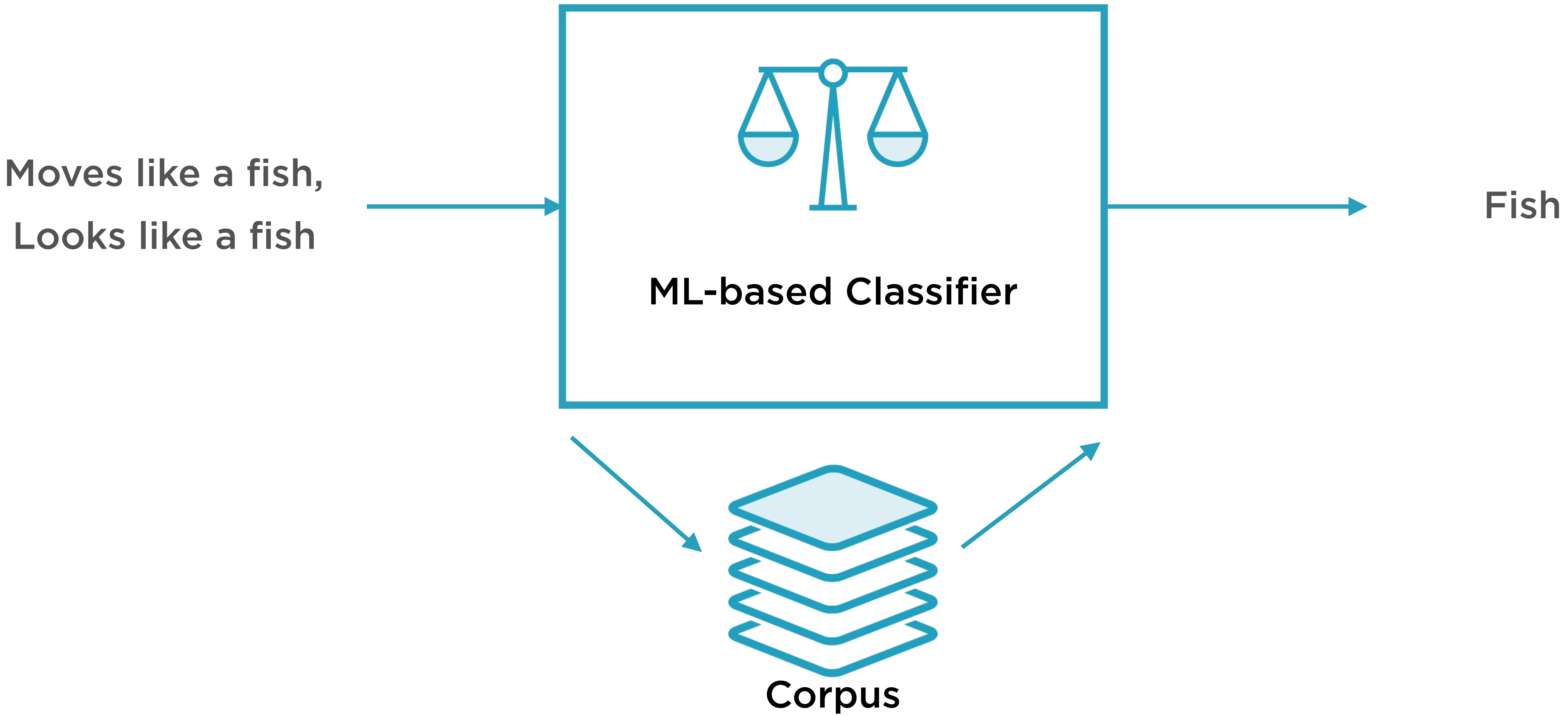
Gives birth like a mammal

Input: Feature Vector

“Traditional” ML-based Binary Classifier



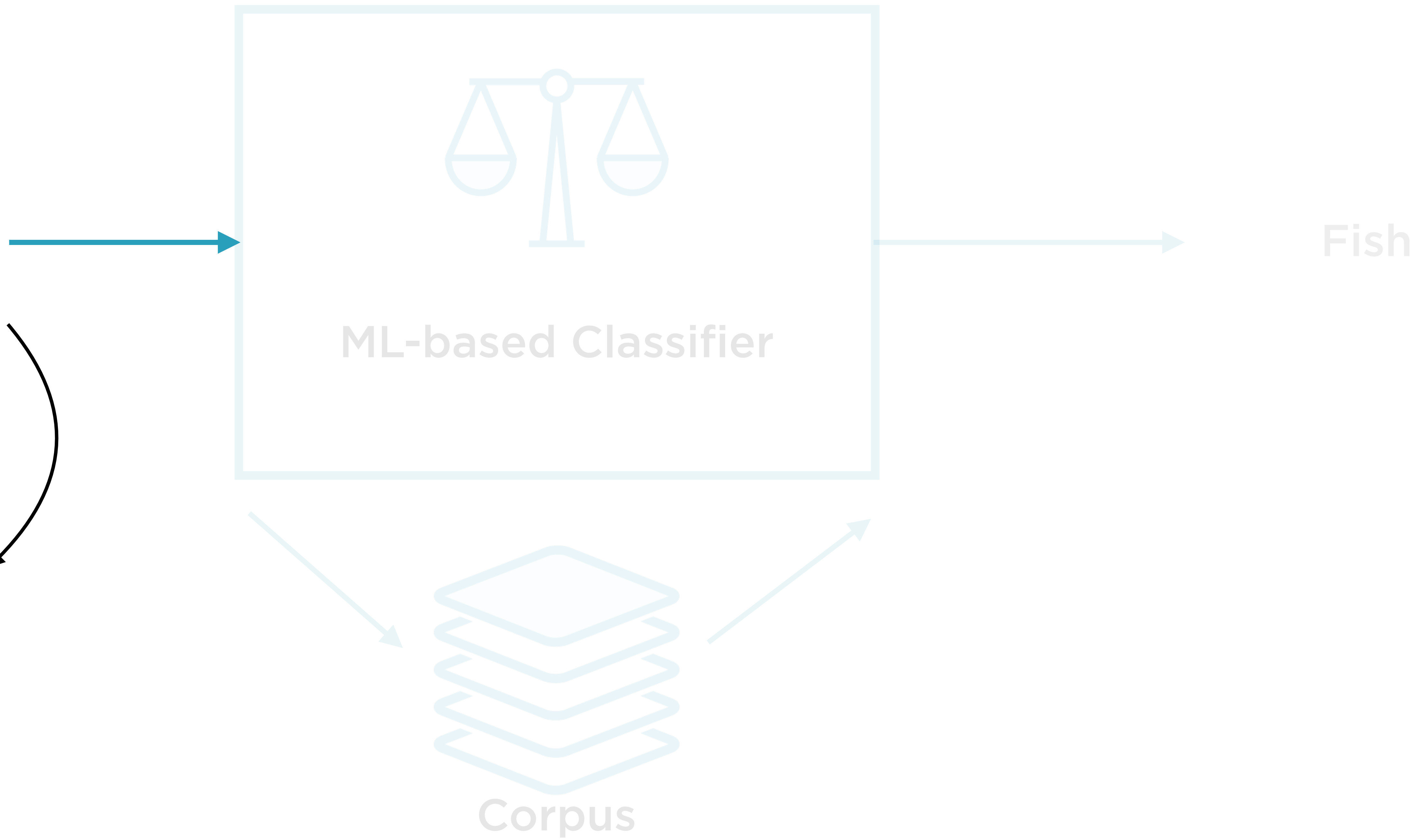
"Traditional" ML-based Binary Classifier



"Traditional" ML-based Binary Classifier

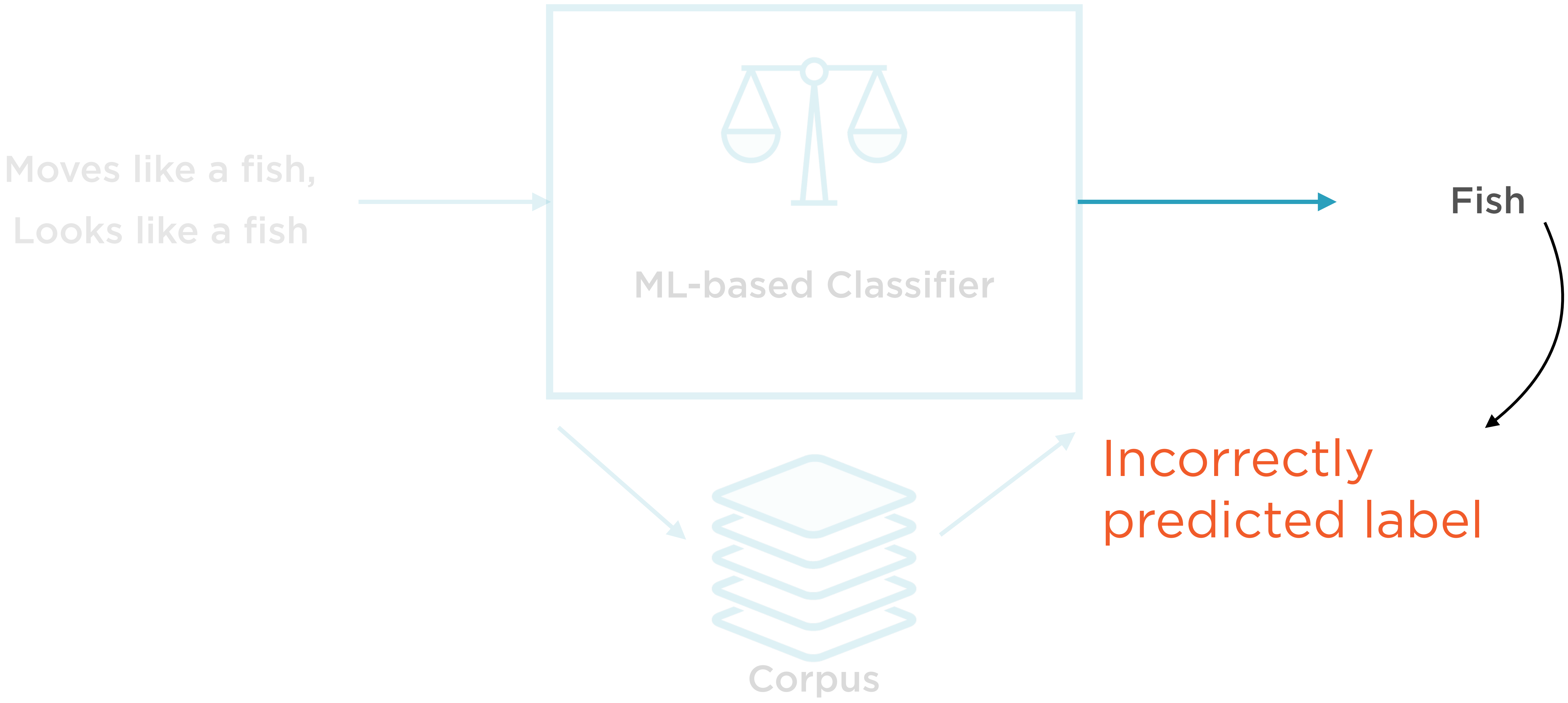
Moves like a fish,

Looks like a fish



Poor choice of features

“Traditional” ML-based Binary Classifier



Feature Vectors

The attributes that the ML algorithm focuses on are called features

Each data point is a list - or vector - of such features

Thus, the input into an ML algorithm is a feature vector

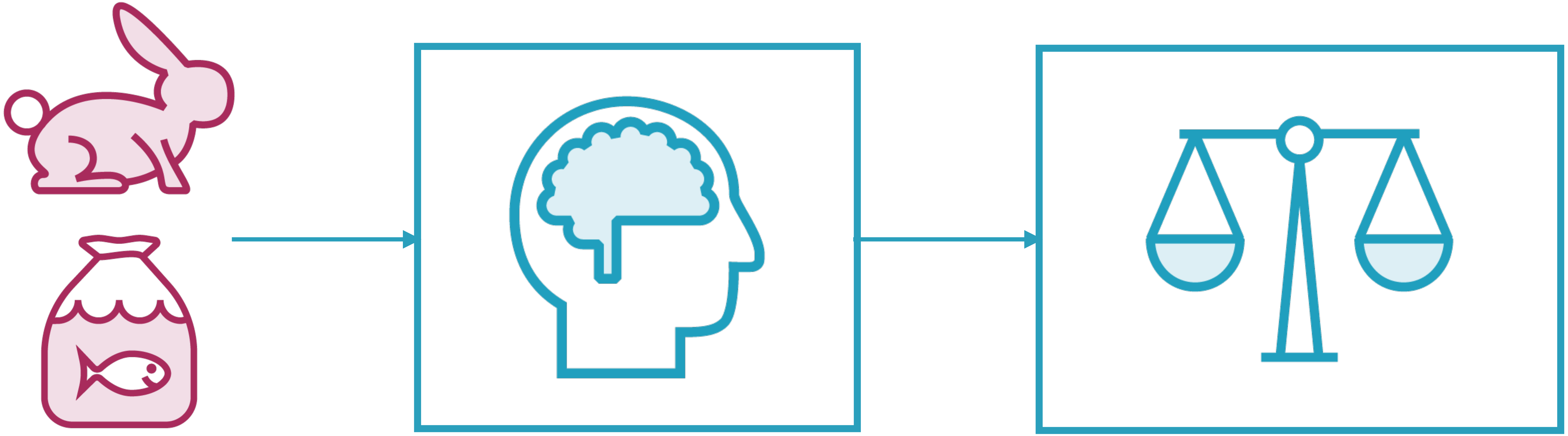
“Traditional” ML-based systems still rely on experts to decide what features to pay attention to

“Representation” ML-based systems figure out by themselves what features to pay attention to

Understanding Deep Learning

“Representation” ML-based systems figure out by themselves what features to pay attention to

“Traditional” ML-based Binary Classifier

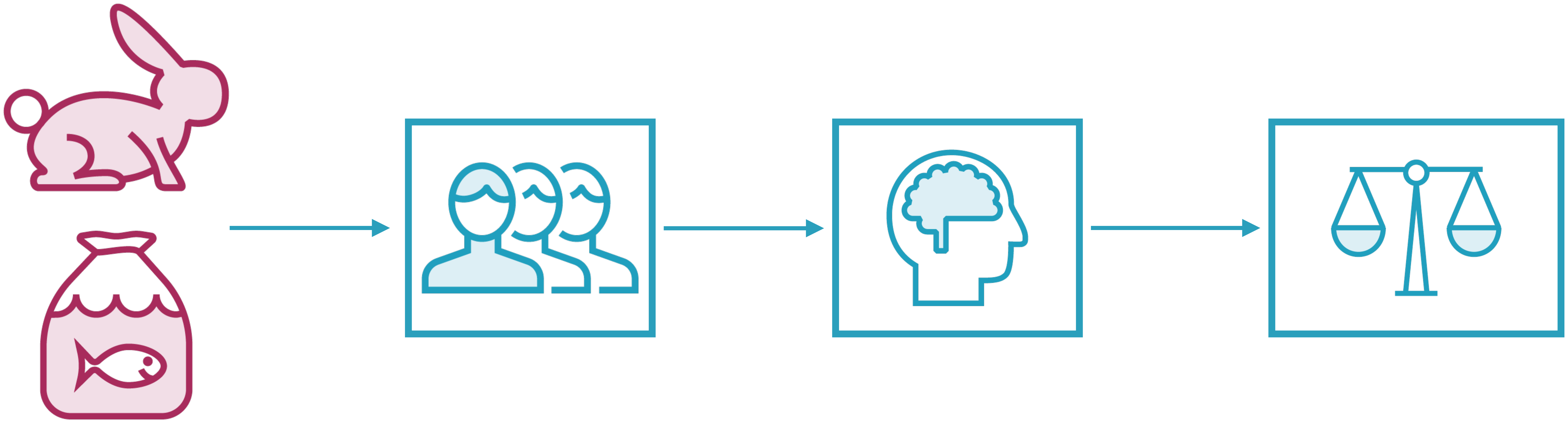


Corpus

Classification Algorithm

ML-based Classifier

“Traditional” ML-based Binary Classifier



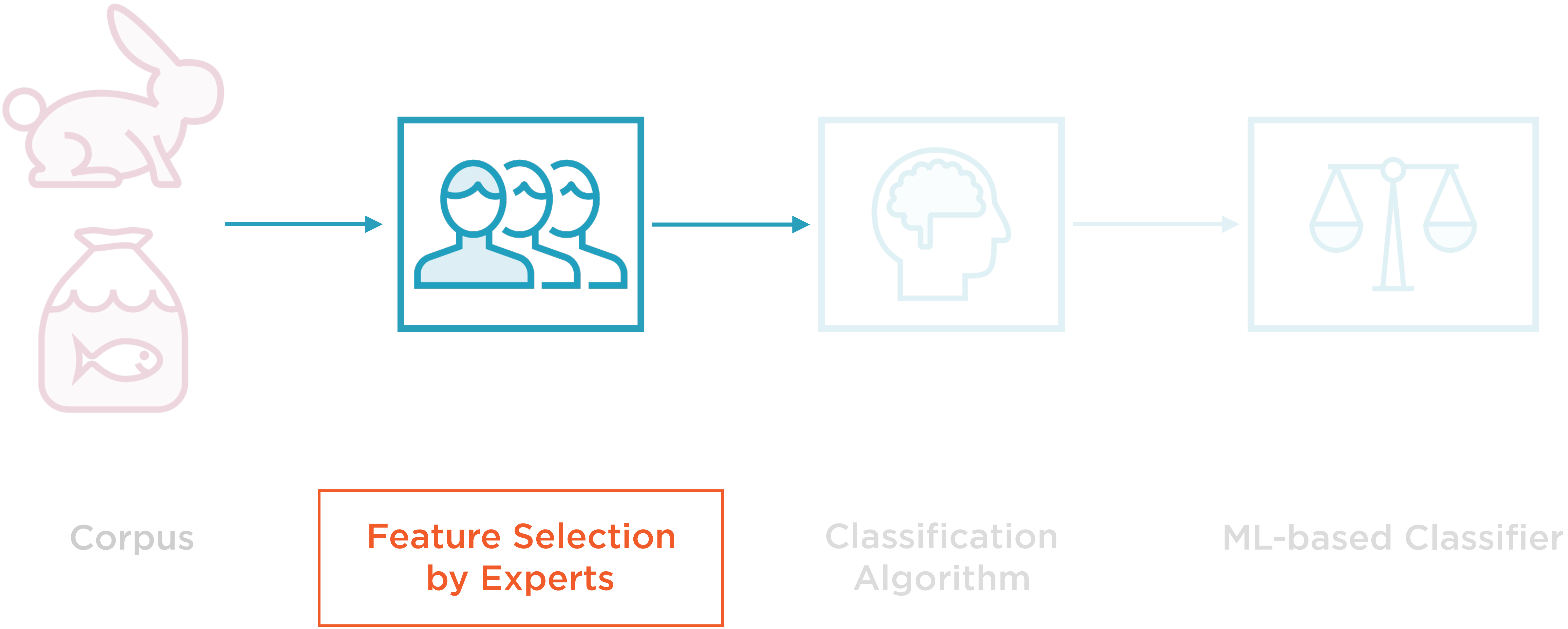
Corpus

Feature Selection
by Experts

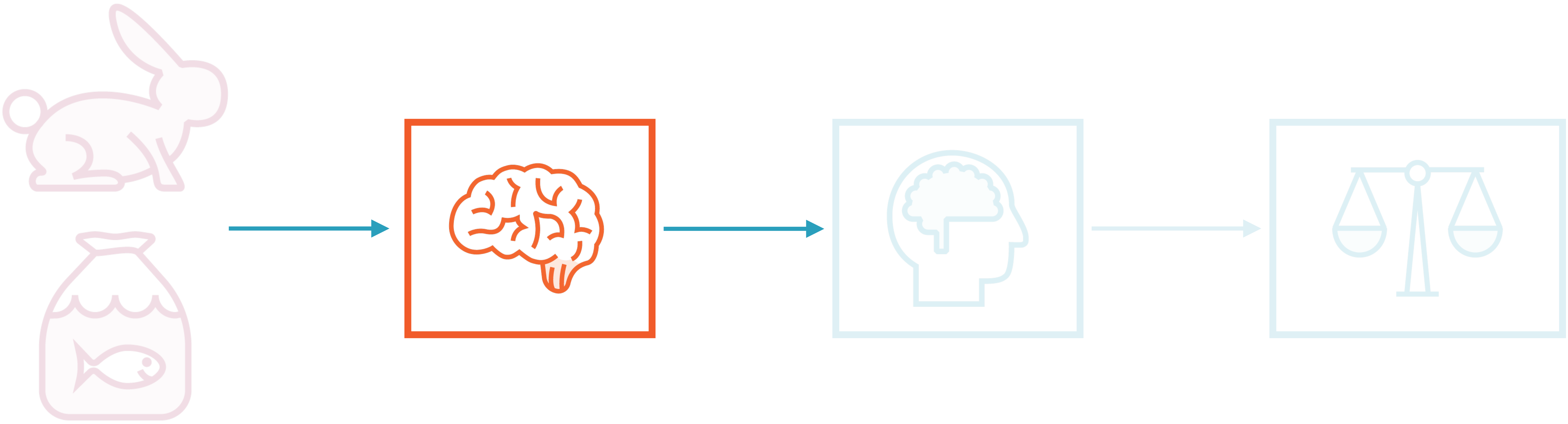
Classification
Algorithm

ML-based Classifier

“Traditional” ML-based Binary Classifier



“Representation” ML-based Binary Classifier



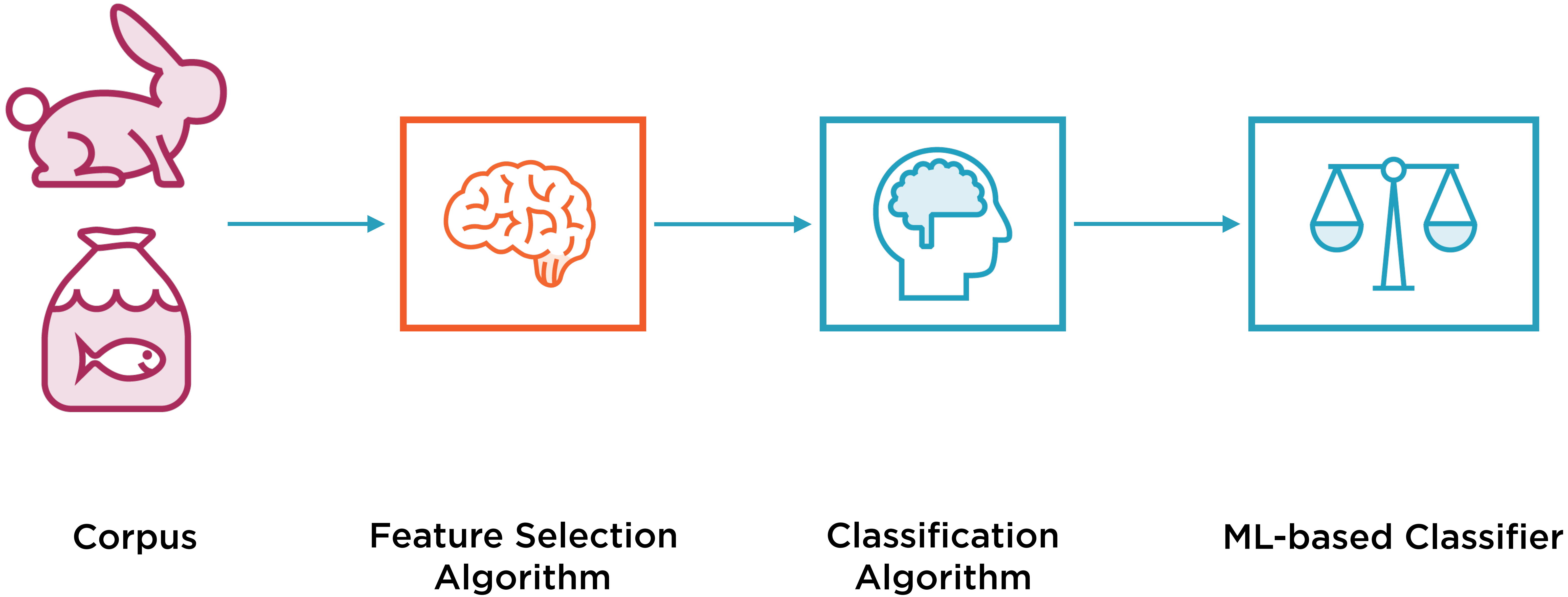
Corpus

Feature Selection Algorithm

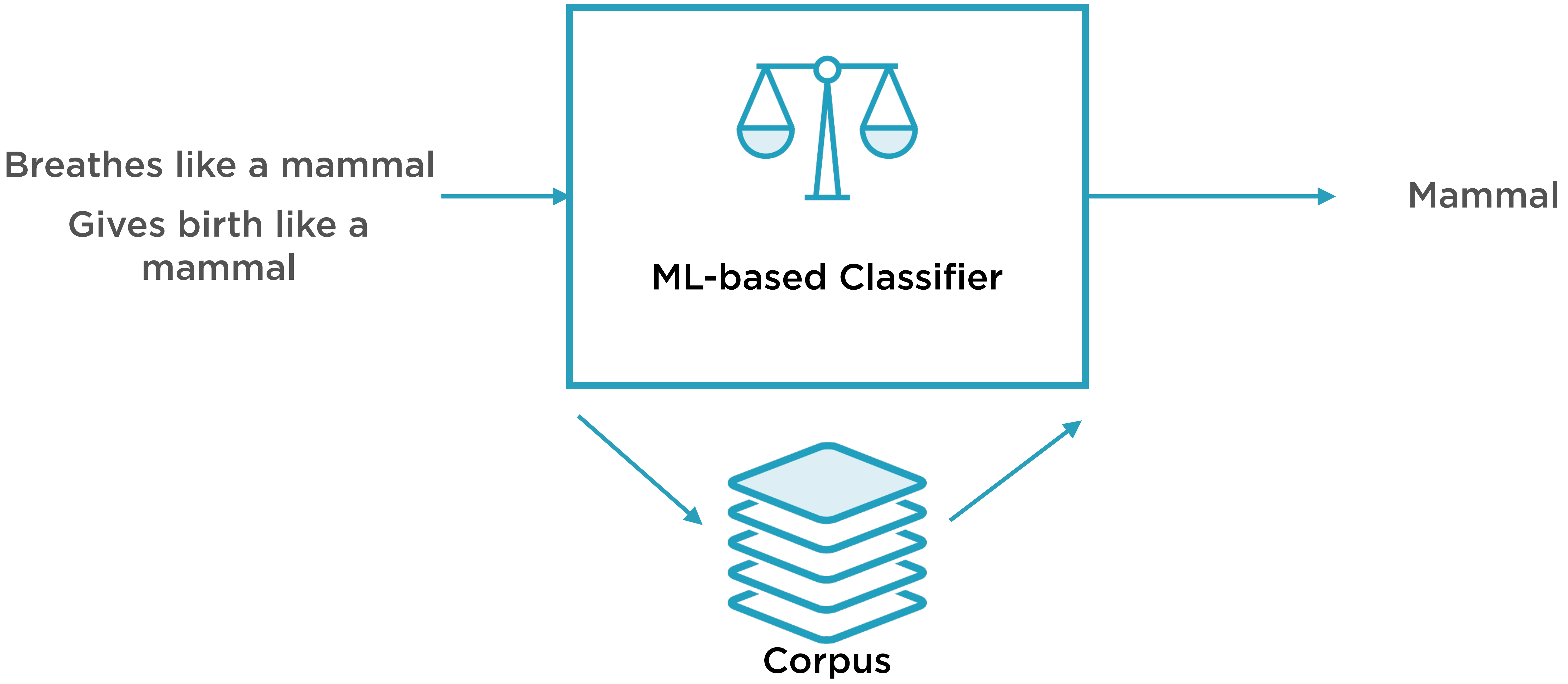
Classification Algorithm

ML-based Classifier

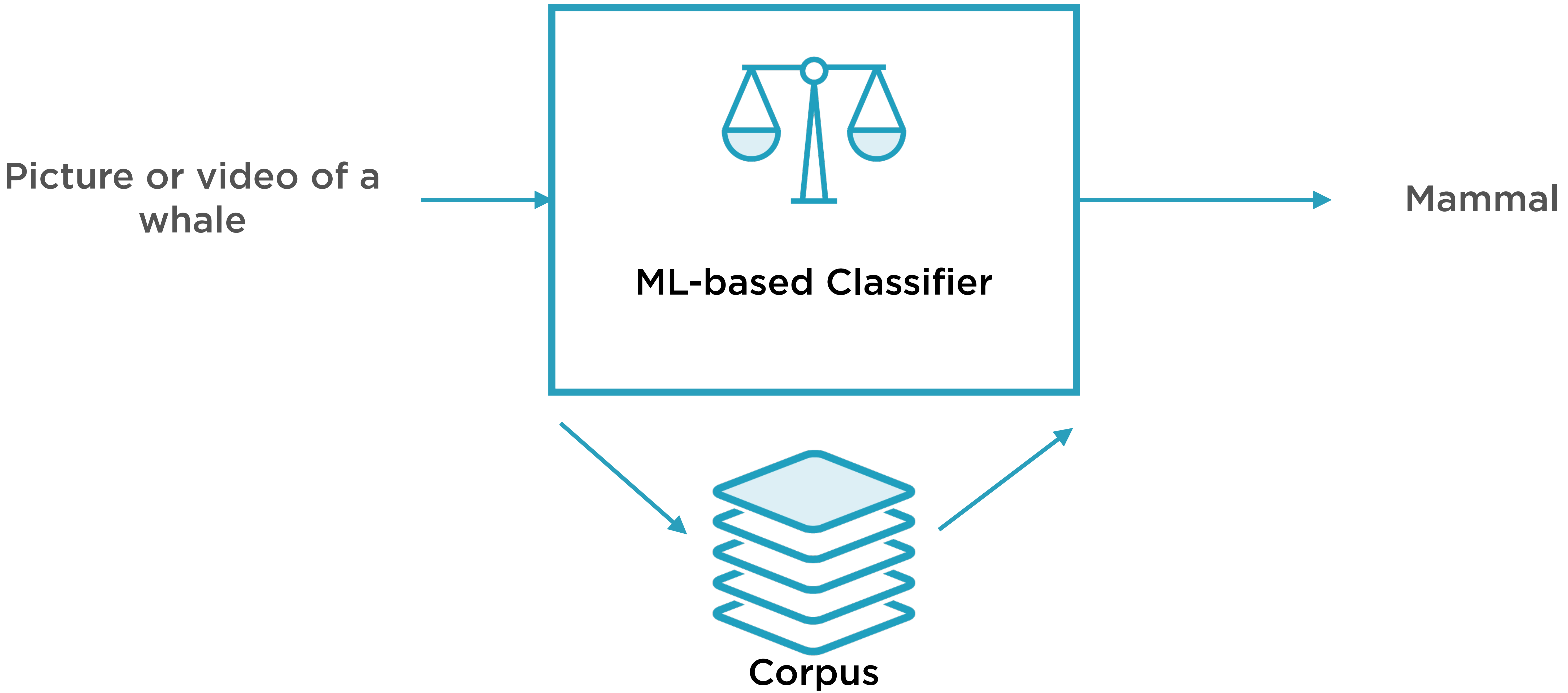
“Representation” ML-based Binary Classifier



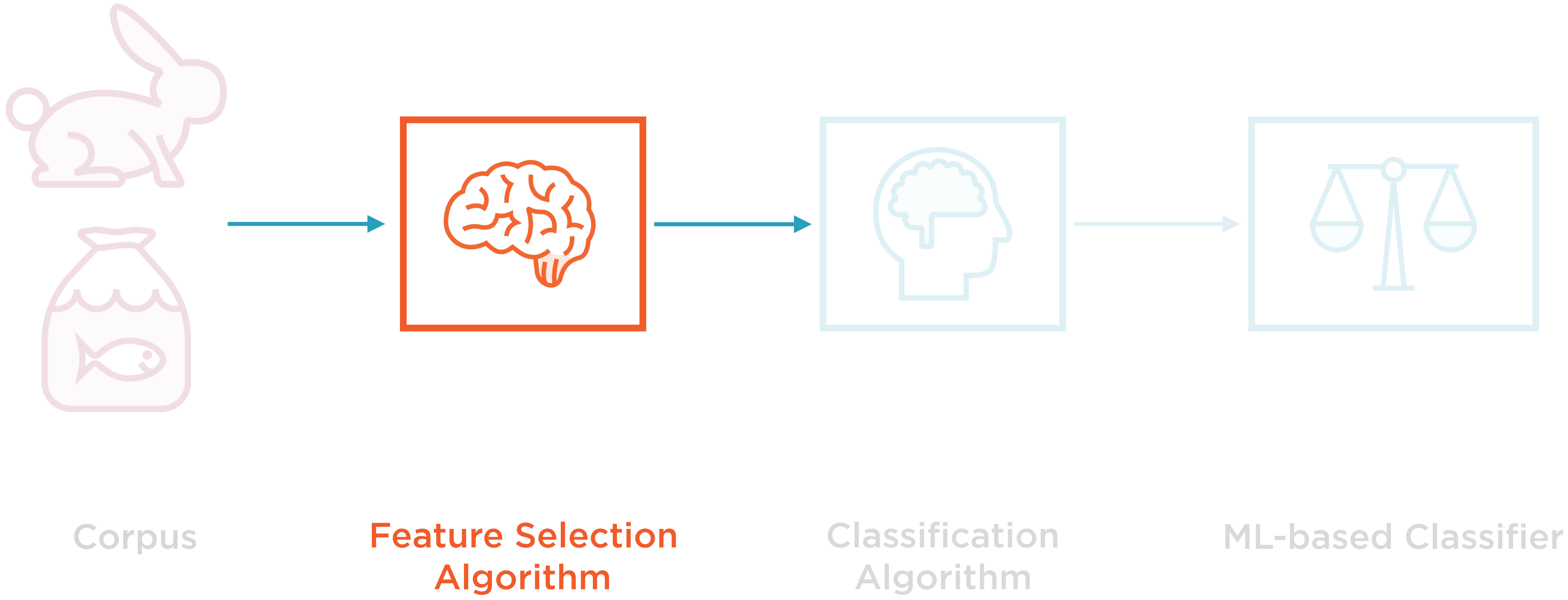
“Traditional” ML-based Binary Classifier



“Representation” ML-based Binary Classifier



“Representation” ML-based Binary Classifier



“Deep Learning” systems are one type of representation systems

Deep Learning and Neural Networks

Deep Learning and Neural Networks

Deep Learning

Algorithms that learn what features matter

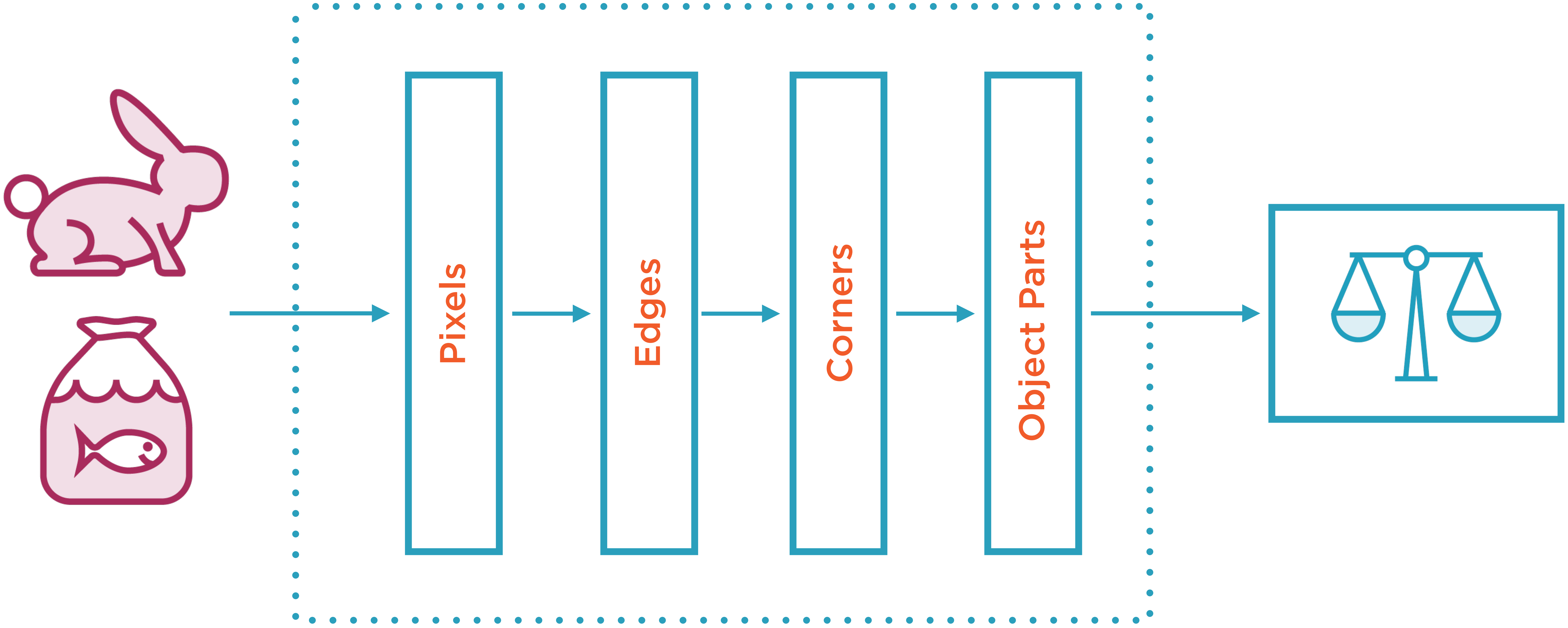
Neural Networks

The most common class of deep learning algorithms

Neurons

Simple building blocks that actually “learn”

“Deep Learning”-based Binary Classifier

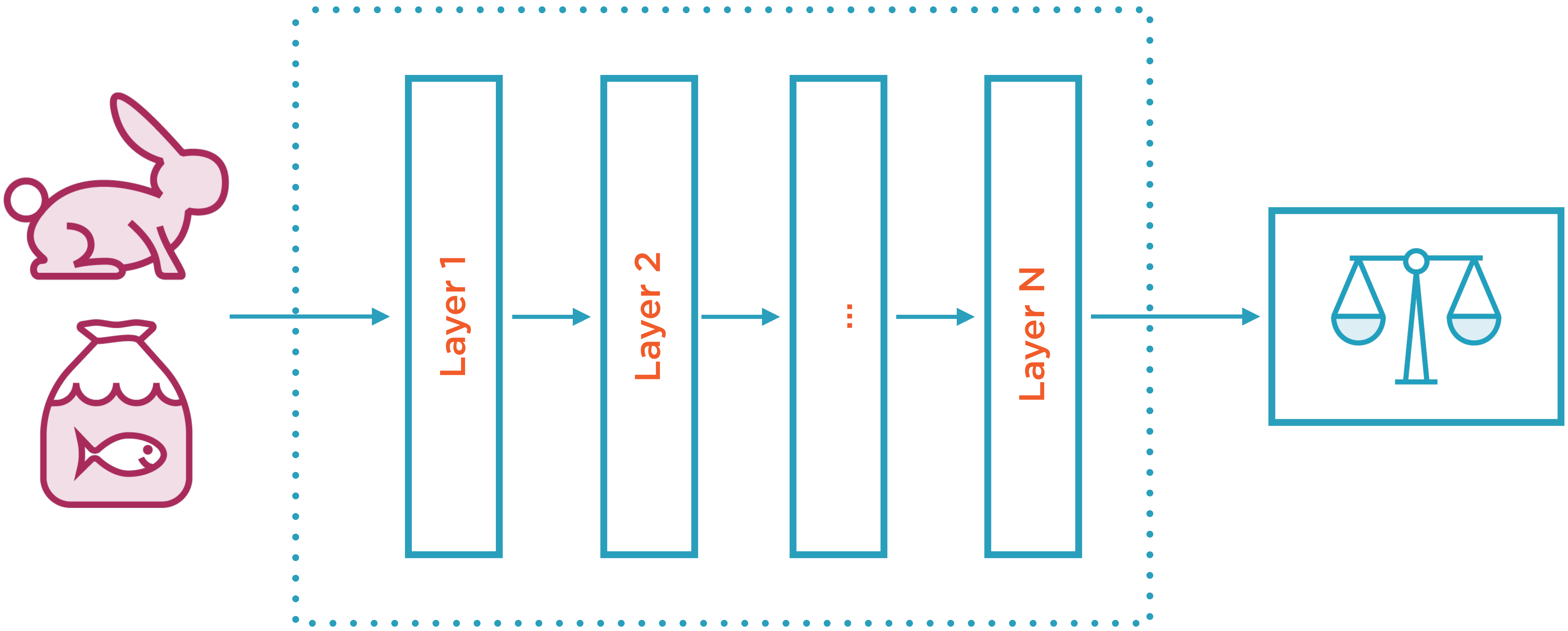


Corpus of Images

Feature Selection & Classification Algorithm

ML-based Classifier

Neural Networks Introduced

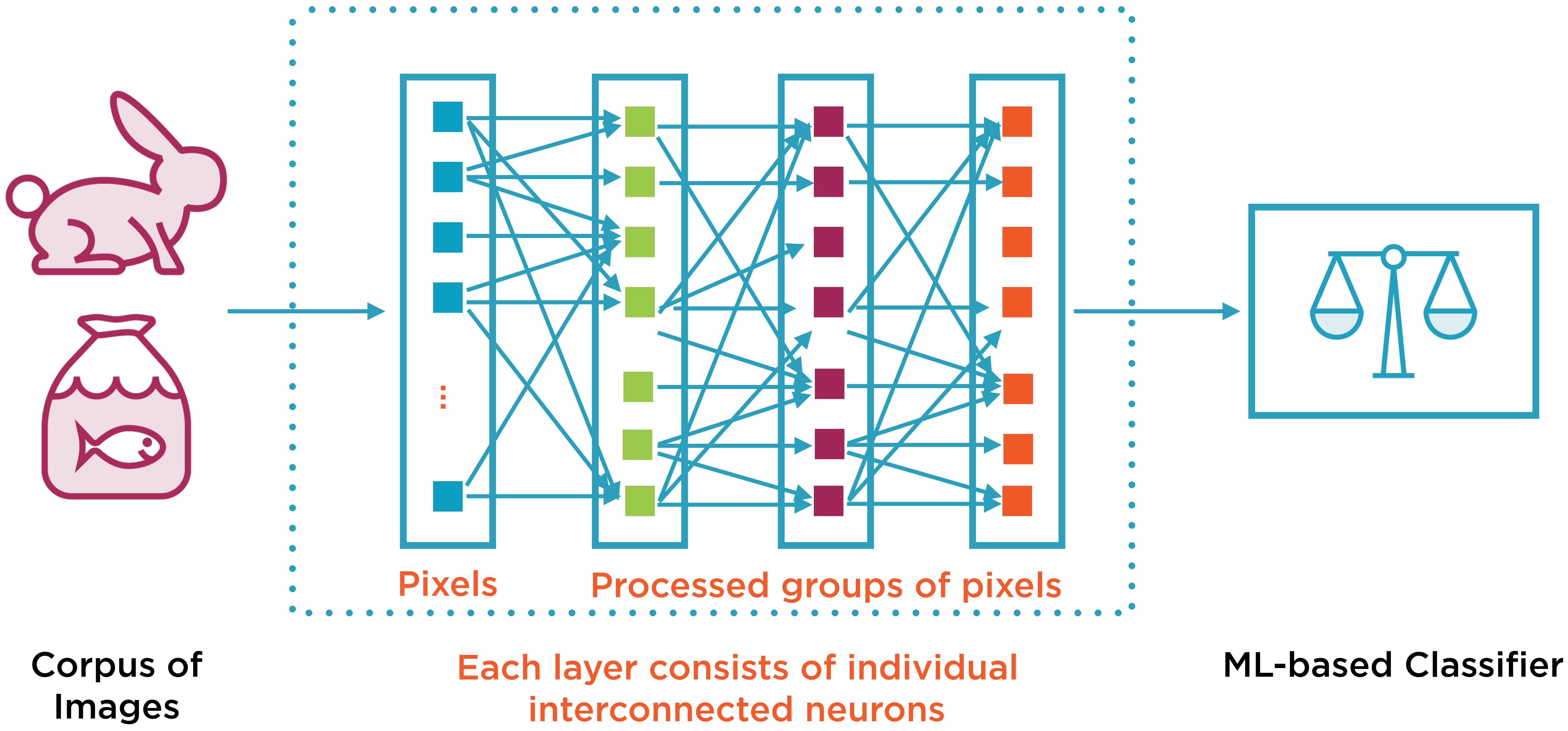


Corpus of Images

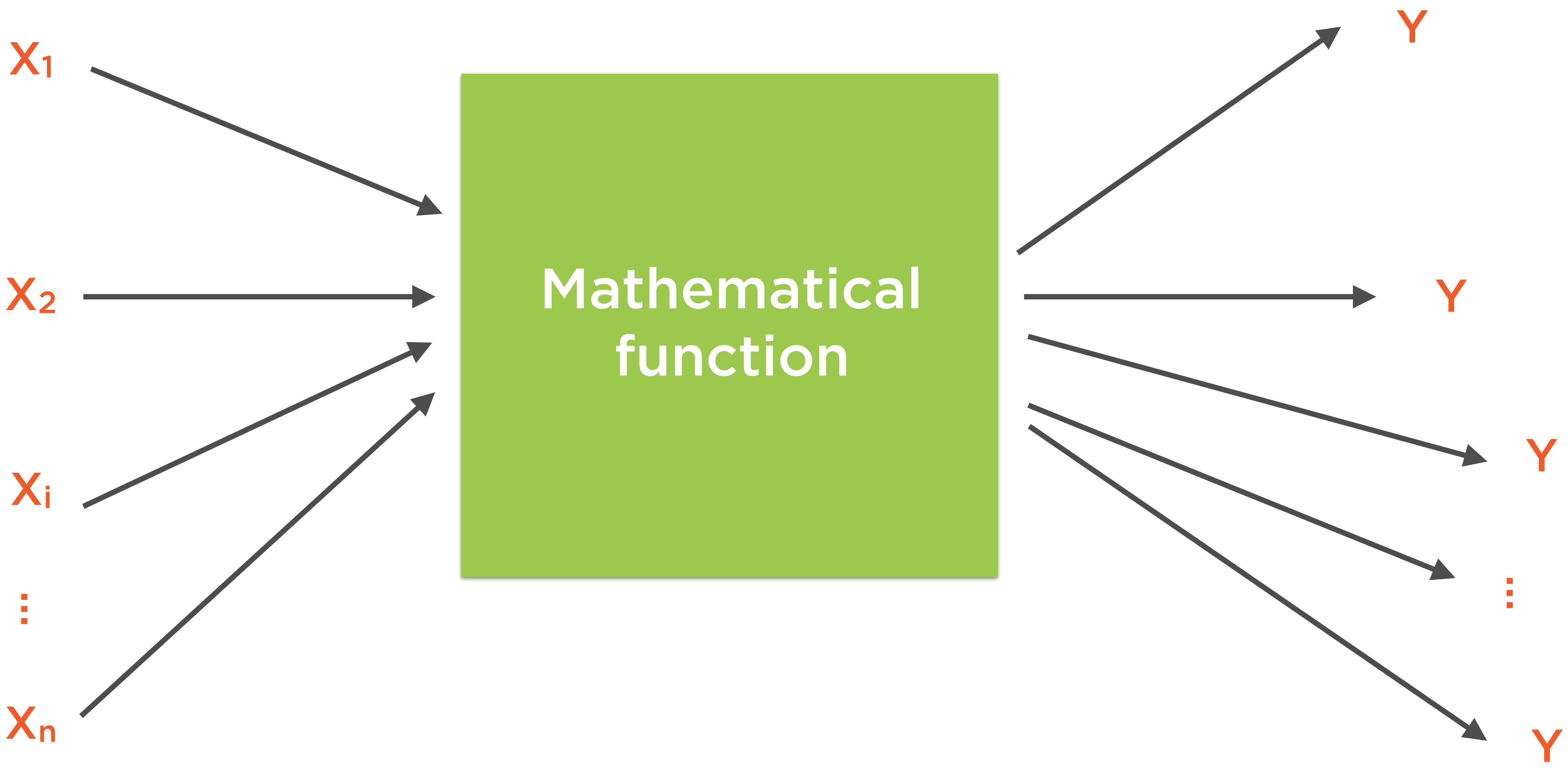
Layers in a neural network

ML-based Classifier

Neural Networks Introduced

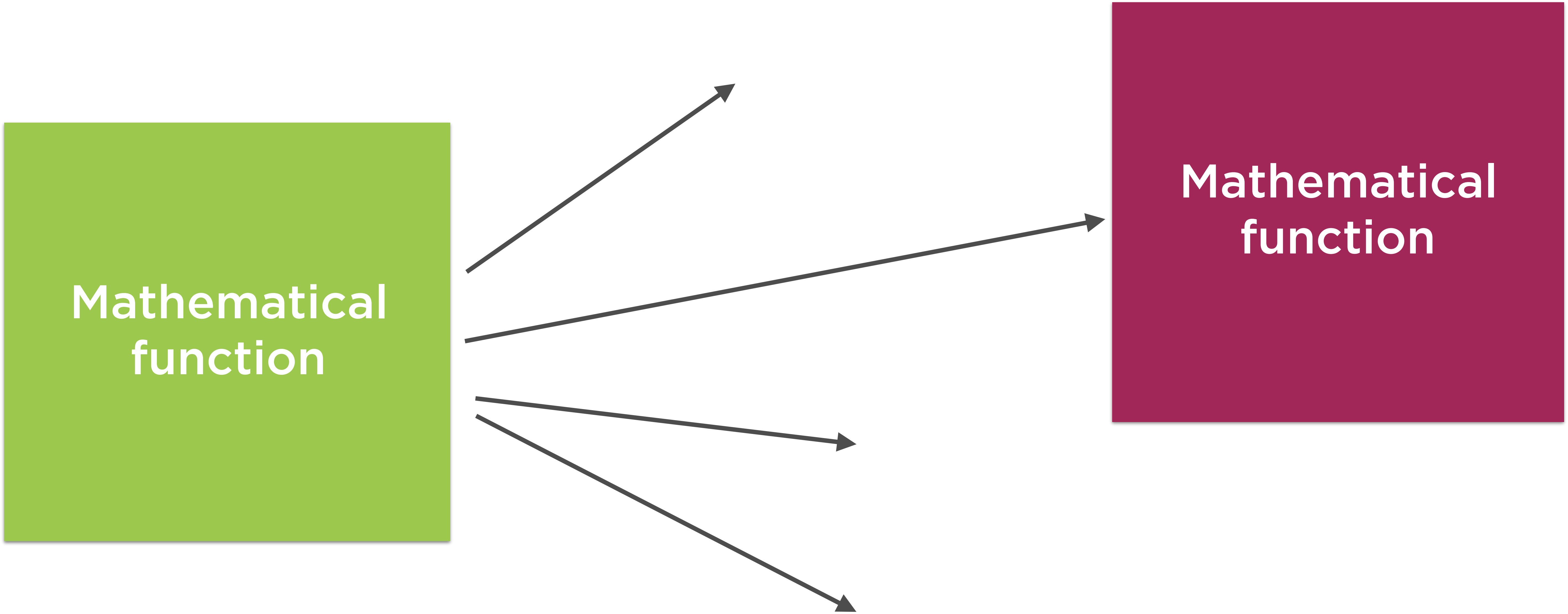


Operation of a Single Neuron



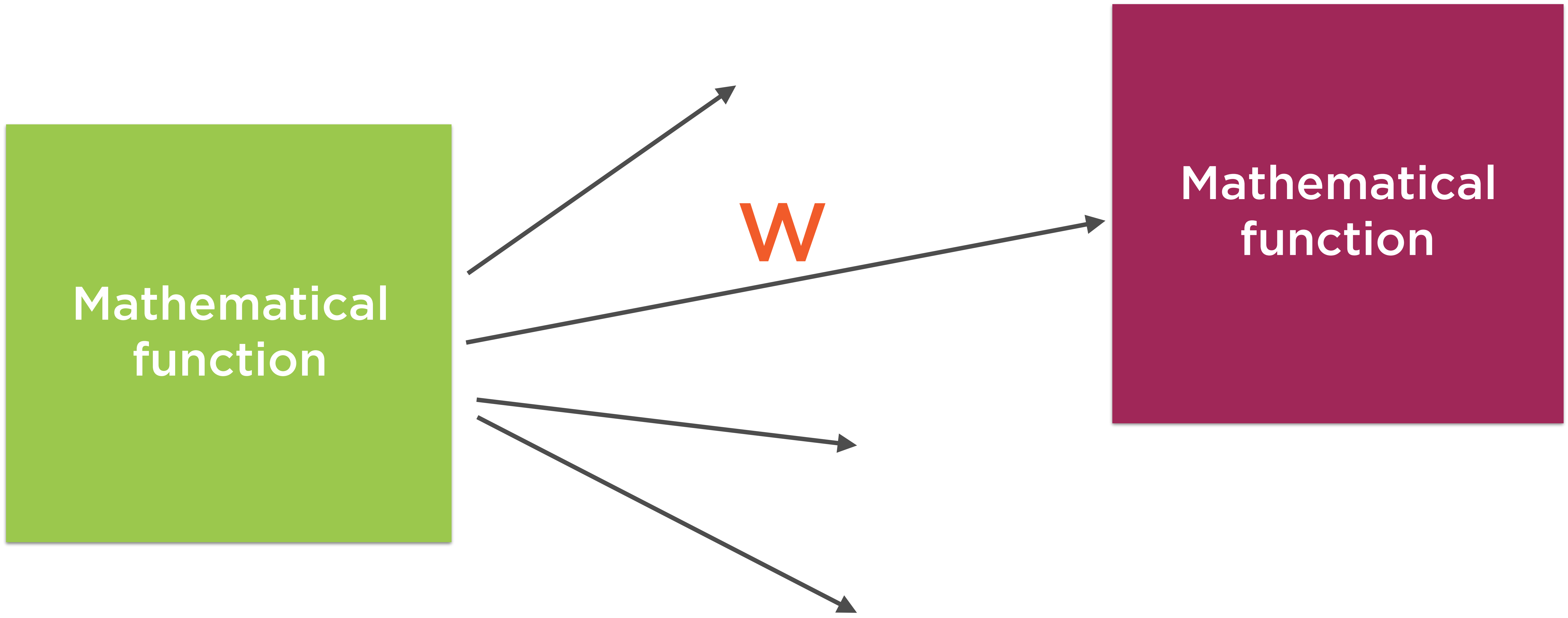
For an active neuron a change in inputs should trigger a corresponding change in the outputs

Operation of a Single Neuron



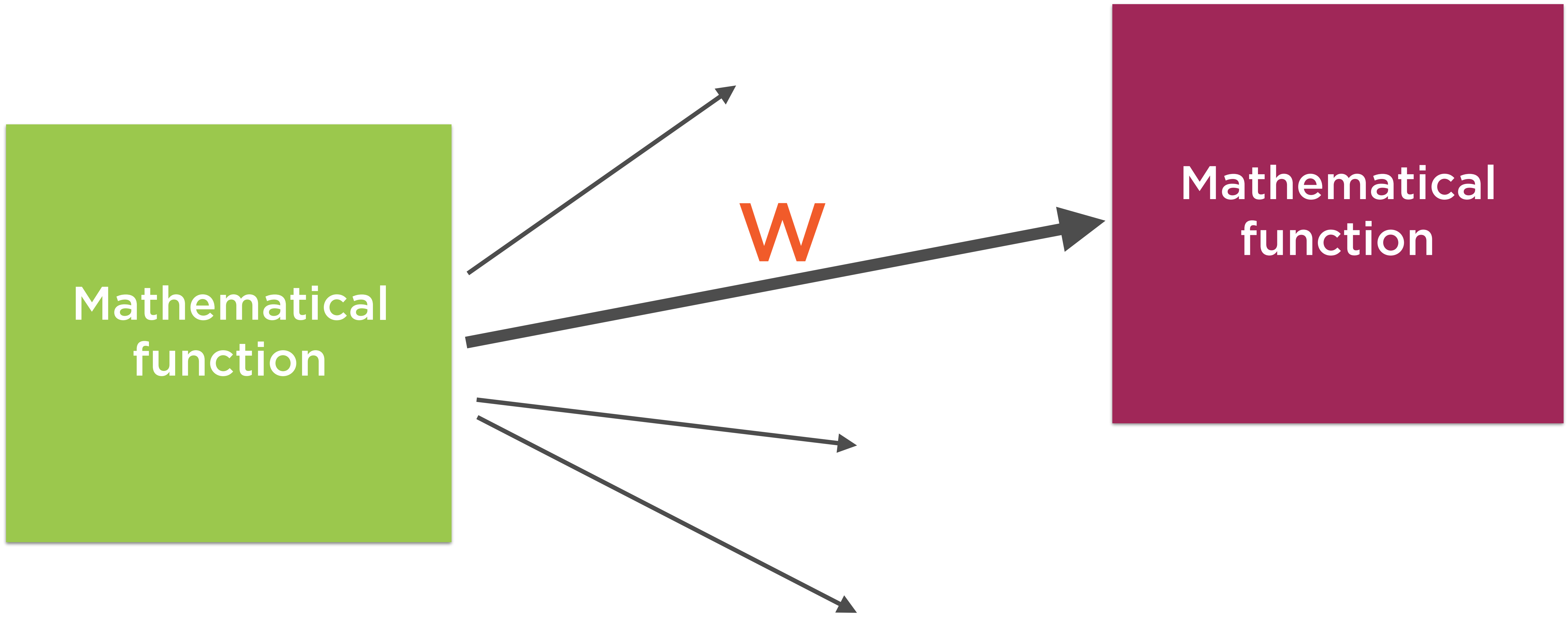
The outputs of neurons feed into the neurons from the next layer

Operation of a Single Neuron



Each connection is associated with a weight

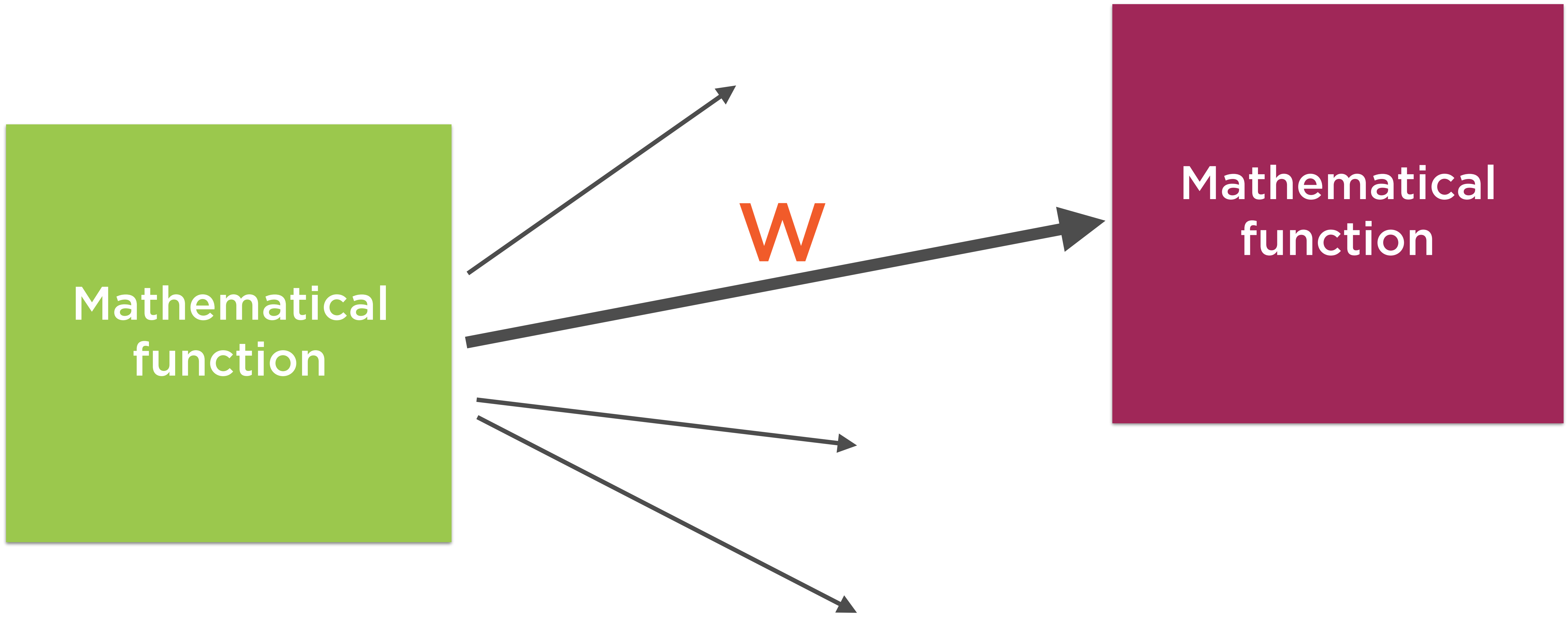
Operation of a Single Neuron



If the second neuron is sensitive to the output of the first neuron, the connection between them gets stronger

W increases

Operation of a Single Neuron



Cells that fire together, wire together

Neural networks help find unknown patterns in massive data sets

Traditional ML vs. Deep Learning Algorithms

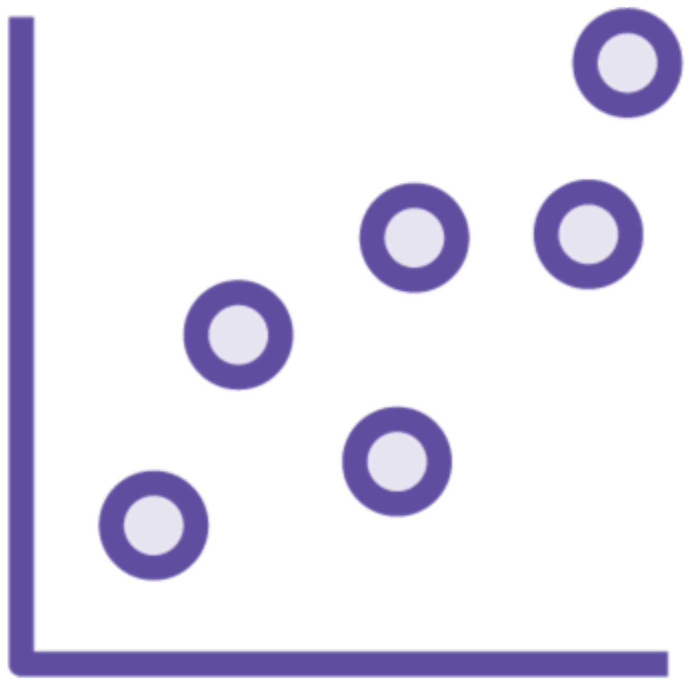
Traditional ML vs. Deep Learning

Traditional ML	Deep Learning
Dynamic	Also dynamic
Relatively little need for expert skill to select features	Even less need for expert skill for feature selection
Experts select features	Algorithms extract features
Works well for numeric data	Also works well for numeric data
Algorithms not specialized to work with images, text	Neural networks at their best dealing with images, videos, complex text

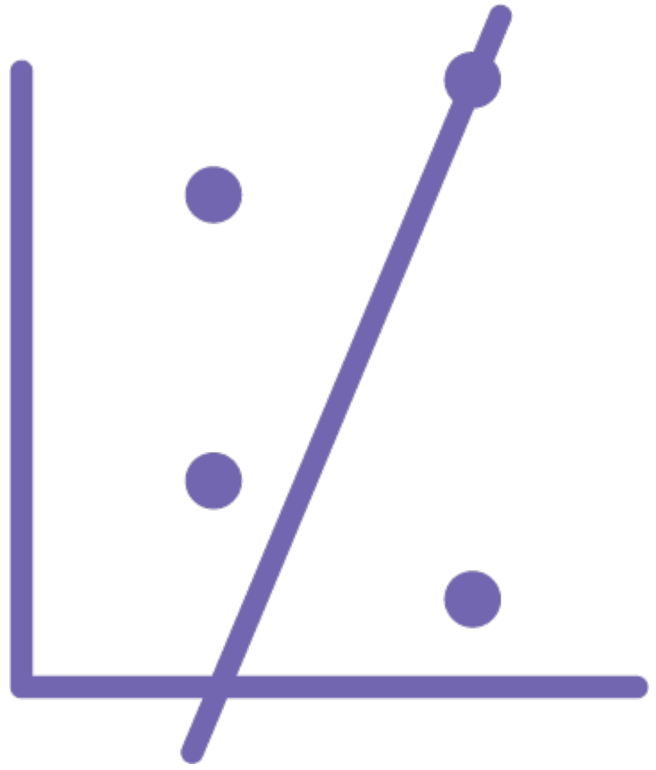
Traditional ML vs. Deep Learning

Traditional ML	Deep Learning
Explicit algorithm	No explicit algorithm - black box
Model a tree structure, find a hyperplane, fit a straight line	Can design highly custom neural networks and interconnections
Training explicitly fits model parameter values	Training implicitly optimizes neural network weights and biases

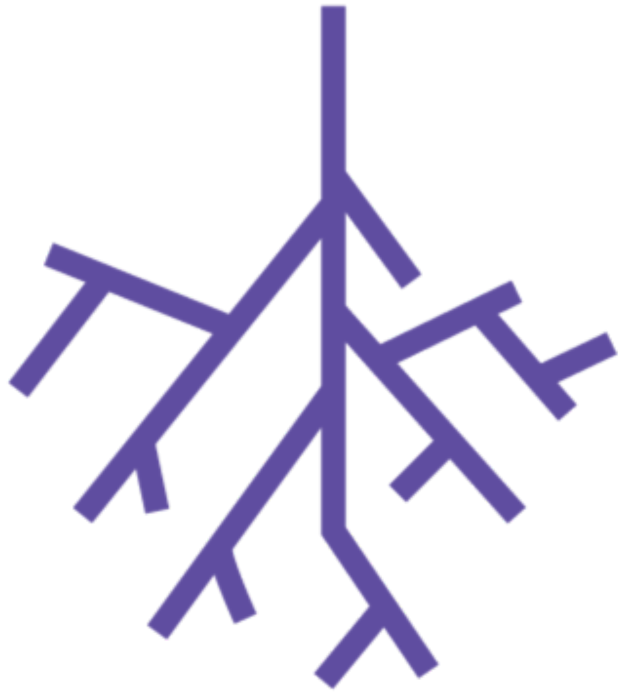
Traditional ML Algorithms



Linear Regression

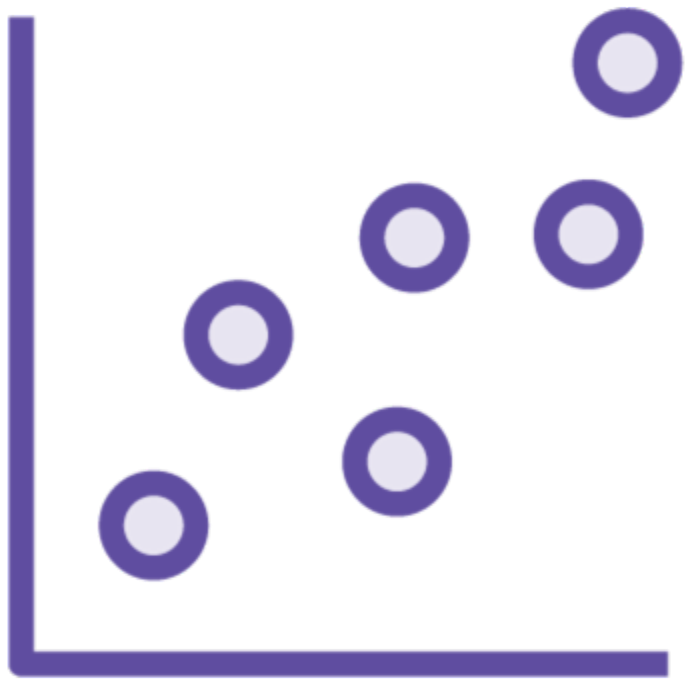


Support Vector Machines



Decision Trees

Traditional ML Algorithms

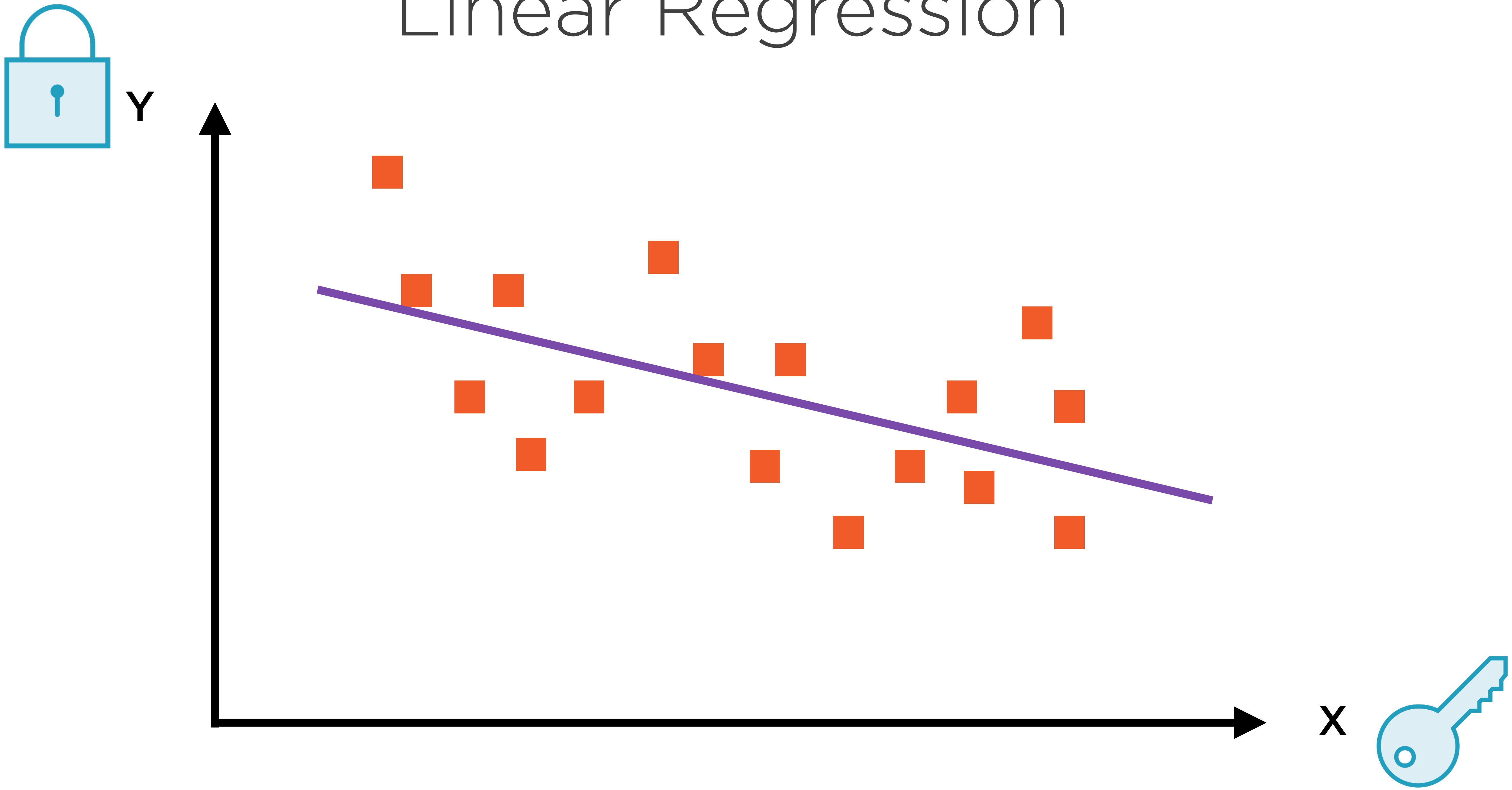


Linear Regression

Support Vector Machines

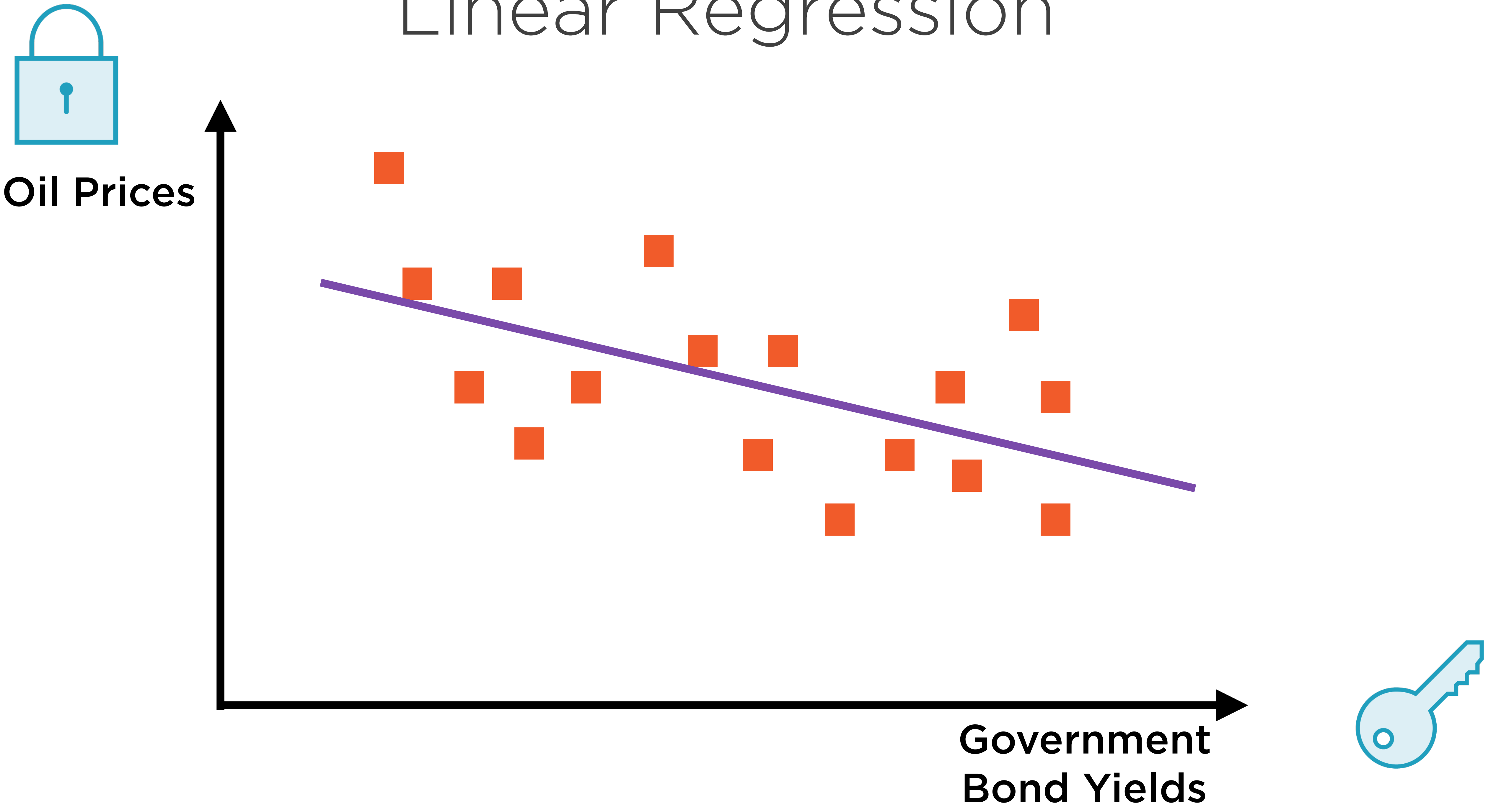
Decision Trees

Linear Regression

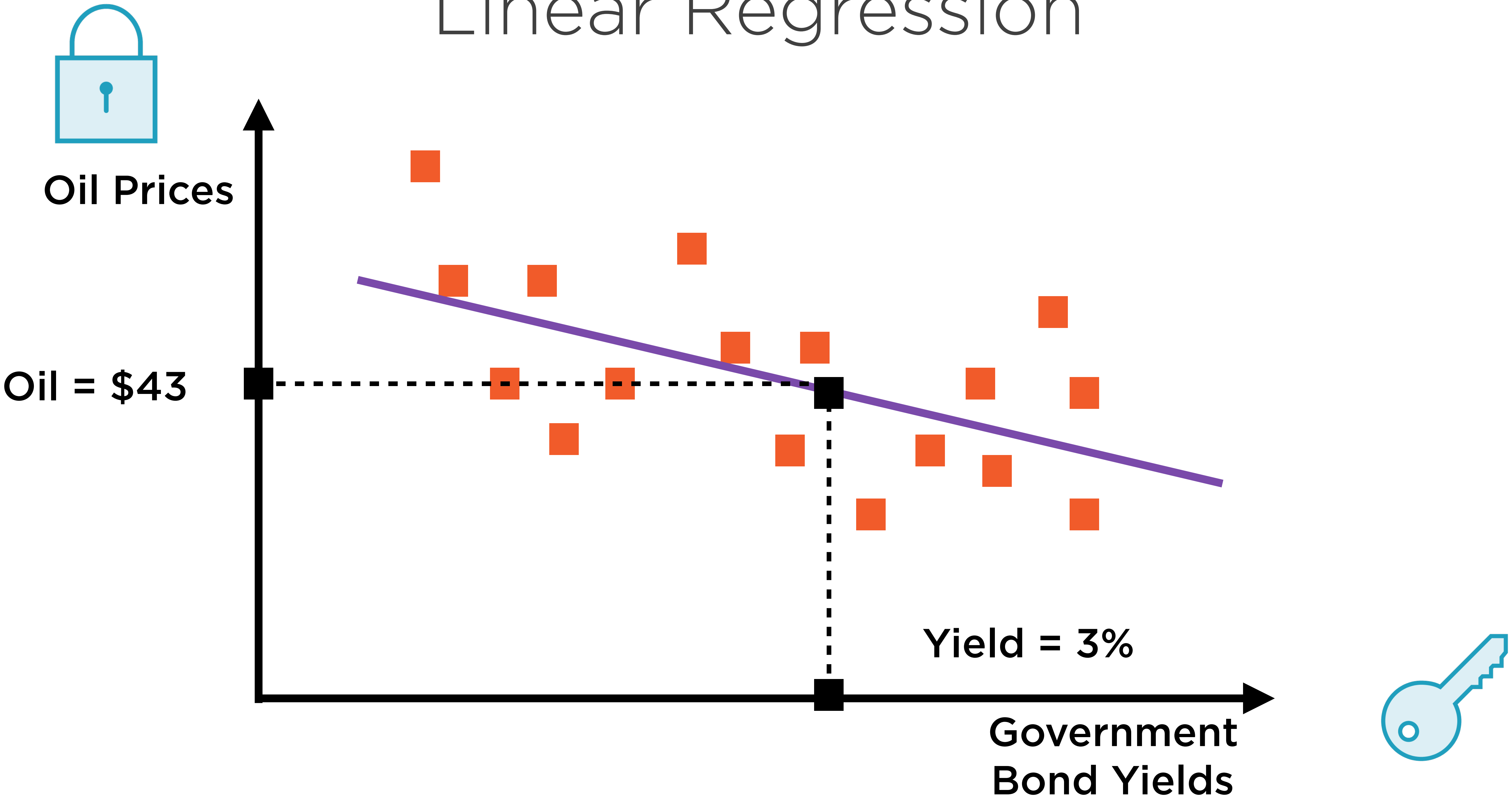


Find the best fit line through the data points

Linear Regression

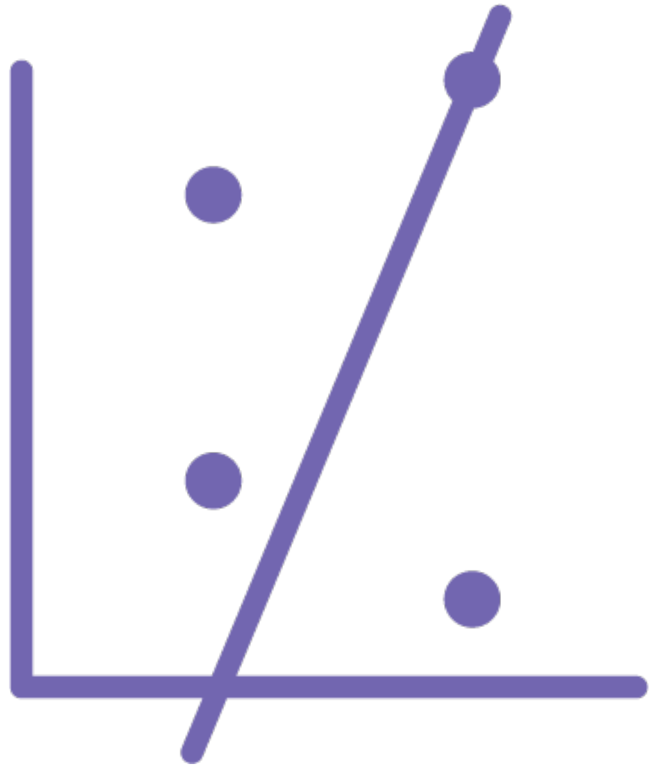


Linear Regression

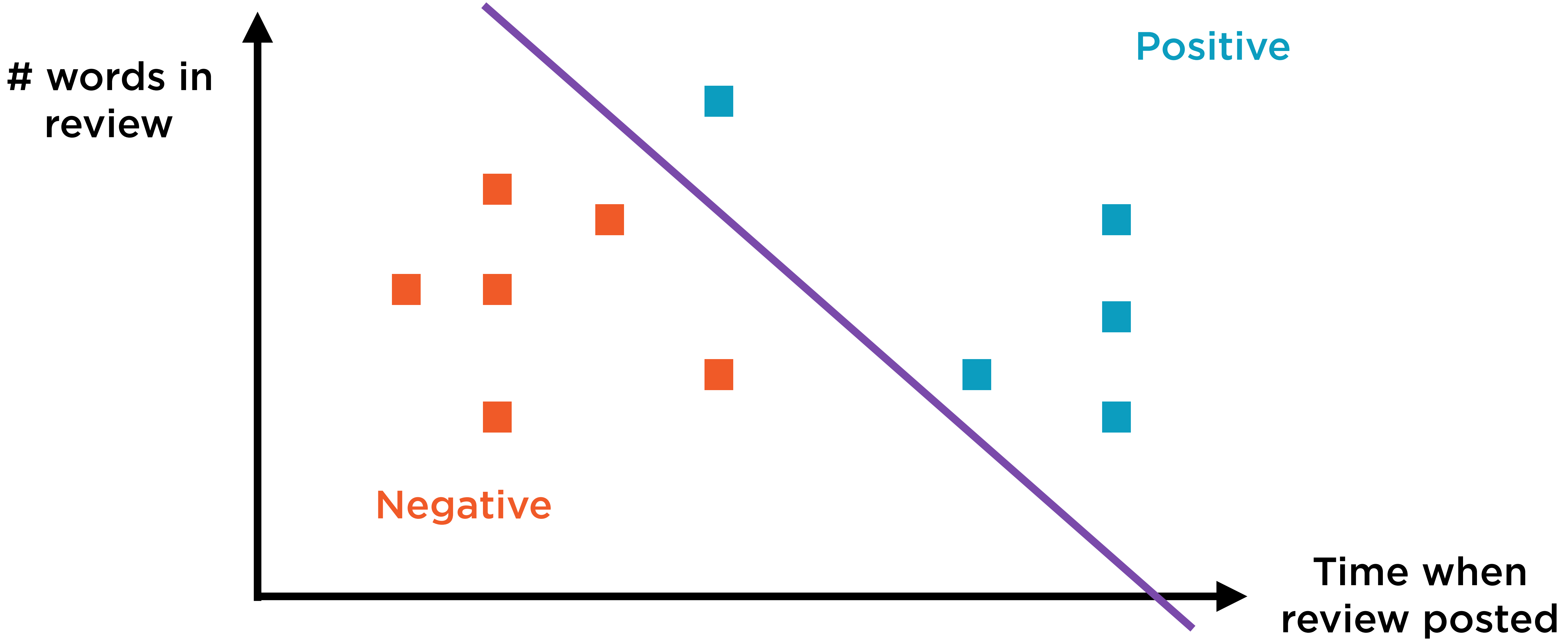


Given a new value of x, use the line to predict the corresponding value of y

Traditional ML Algorithms

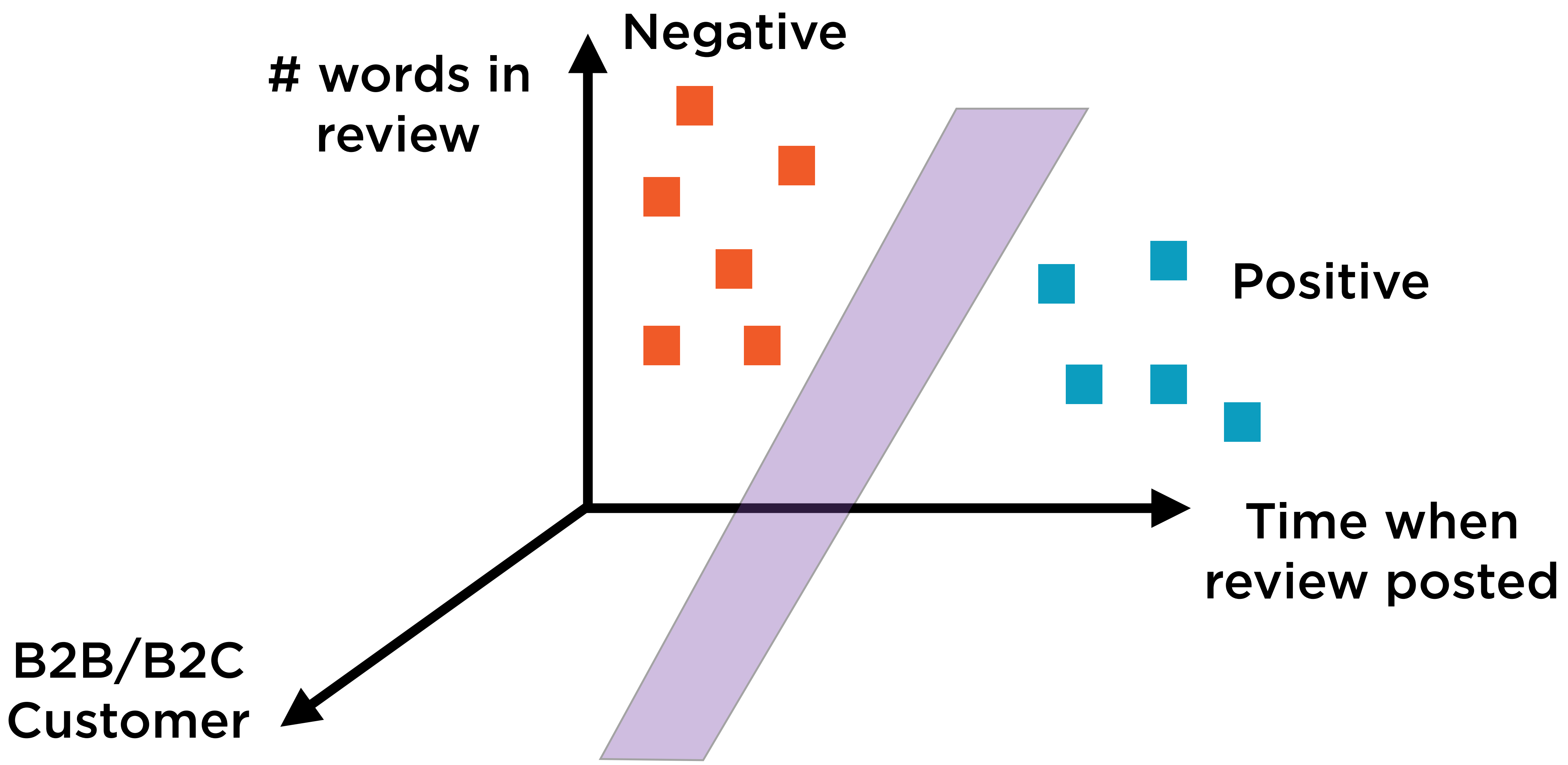


Data in Two Dimensions



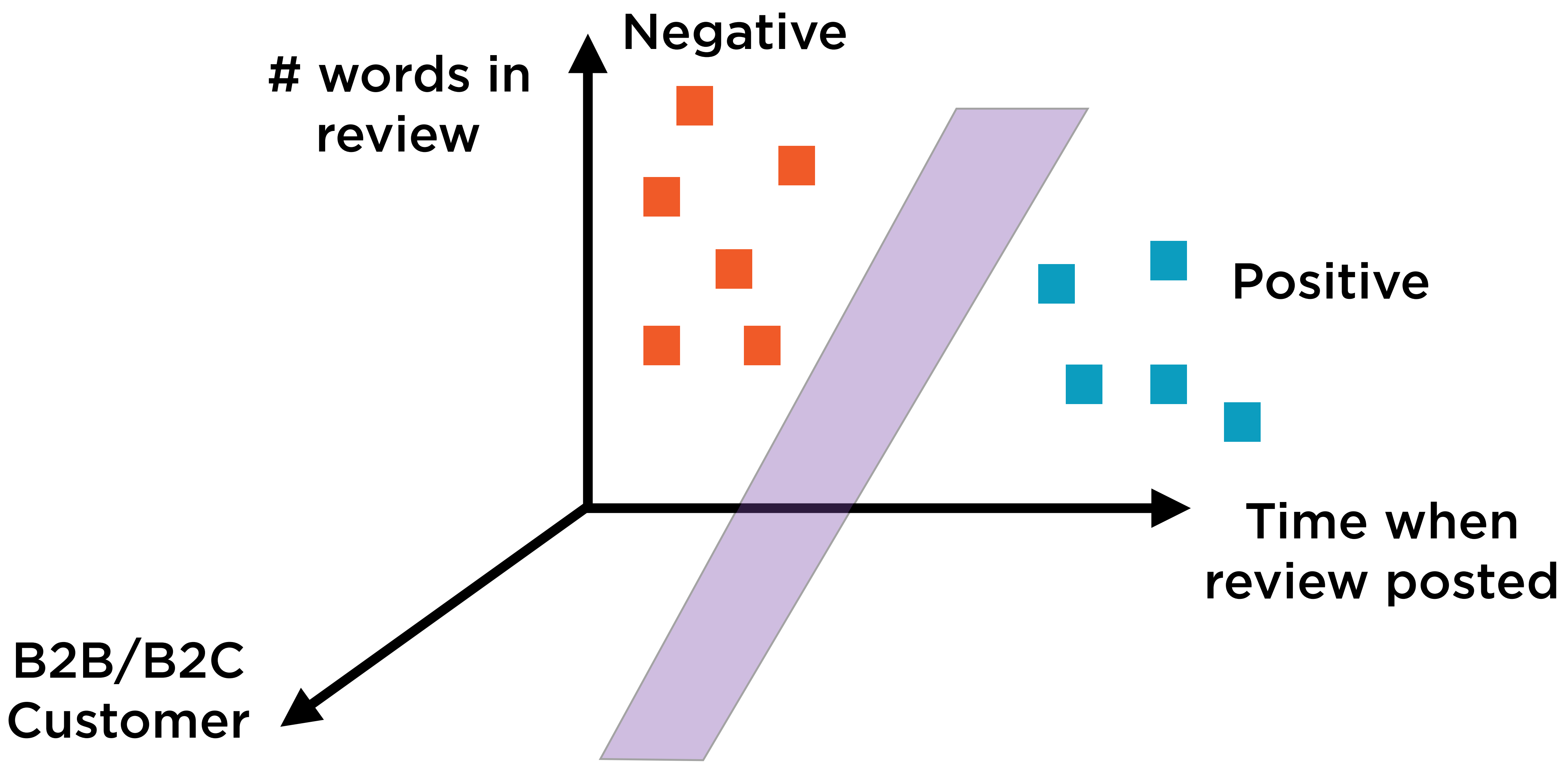
Bidimensional data points can be represented using a plane, and classified using a line

Data in N Dimensions



N-dimensional data can be represented in a hypercube, and classified using a hyperplane

Support Vector Machines

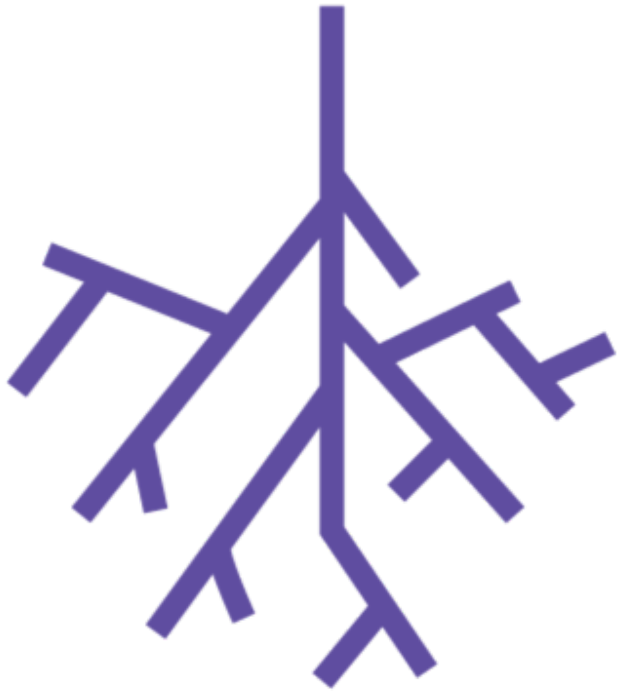


SVM classifiers find the hyperplane that best separates points in a hypercube

Traditional ML Algorithms

Linear Regression

Support Vector Machines



Decision Trees

Jockey or Basketball Player?

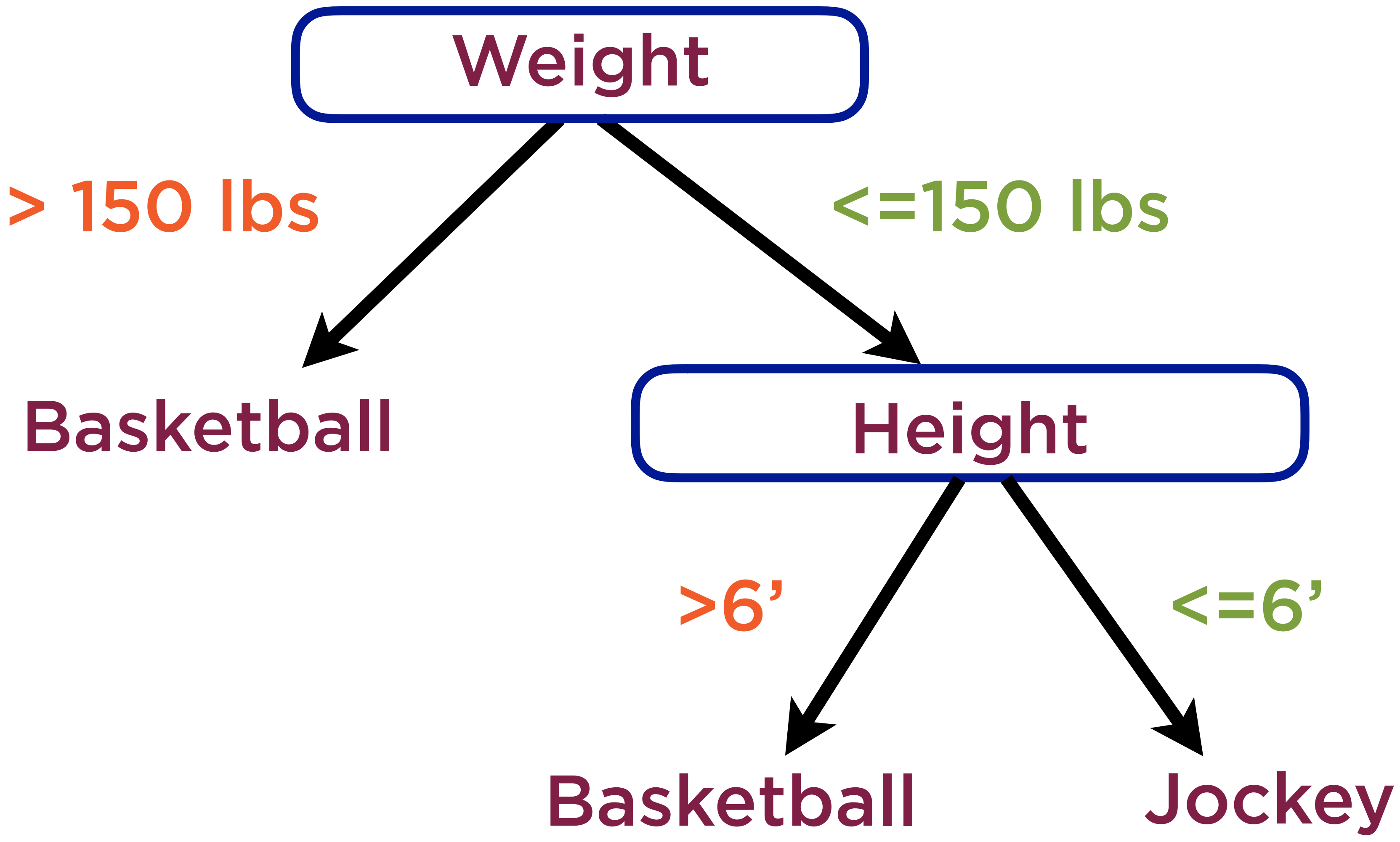
Jockeys

Tend to be light to meet horse carrying limits

Basketball Players

Tend to be tall, strong and heavy

Fit Knowledge Into Rules

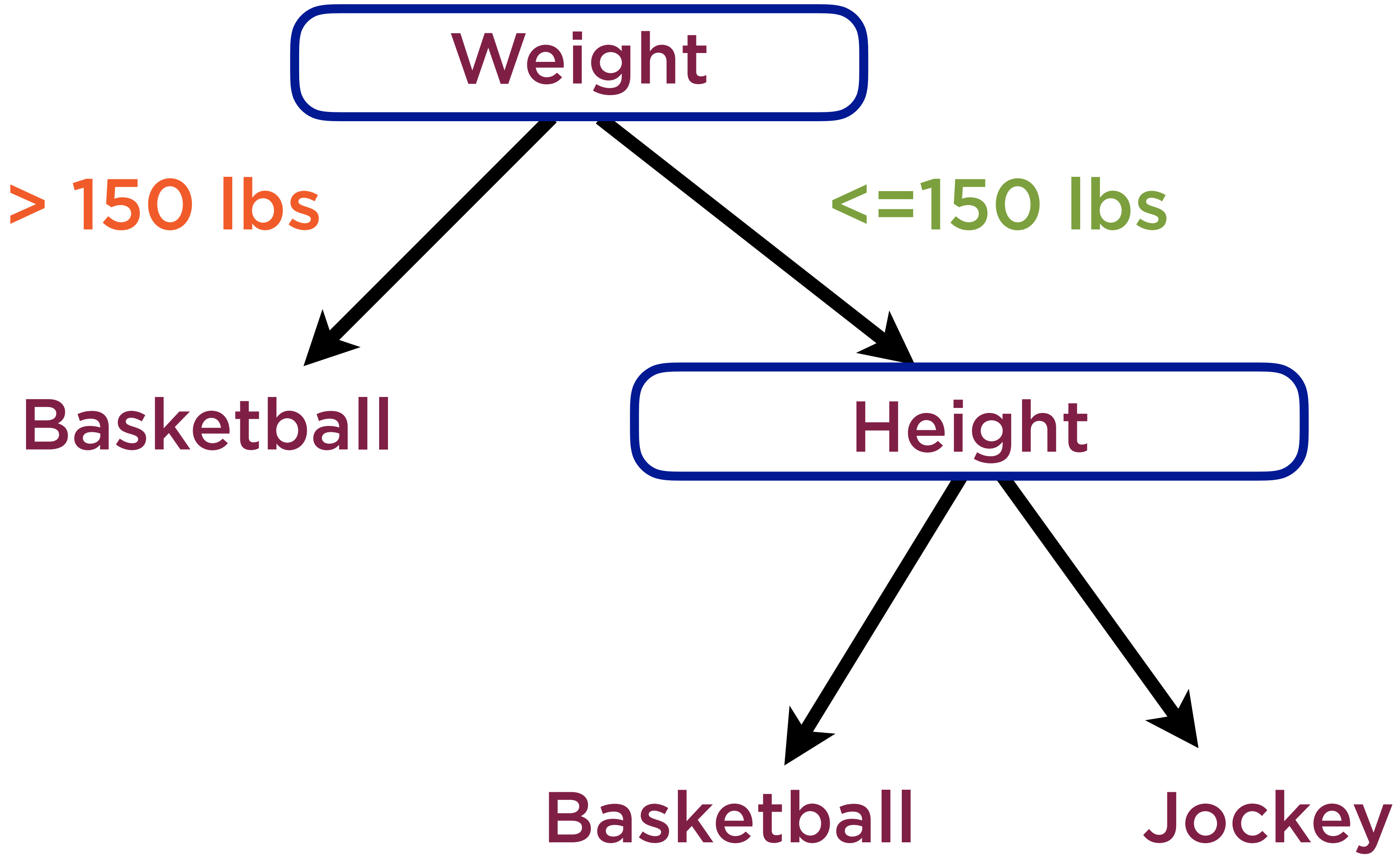


Decision Tree

Fit knowledge into rules

Each rule involves a threshold

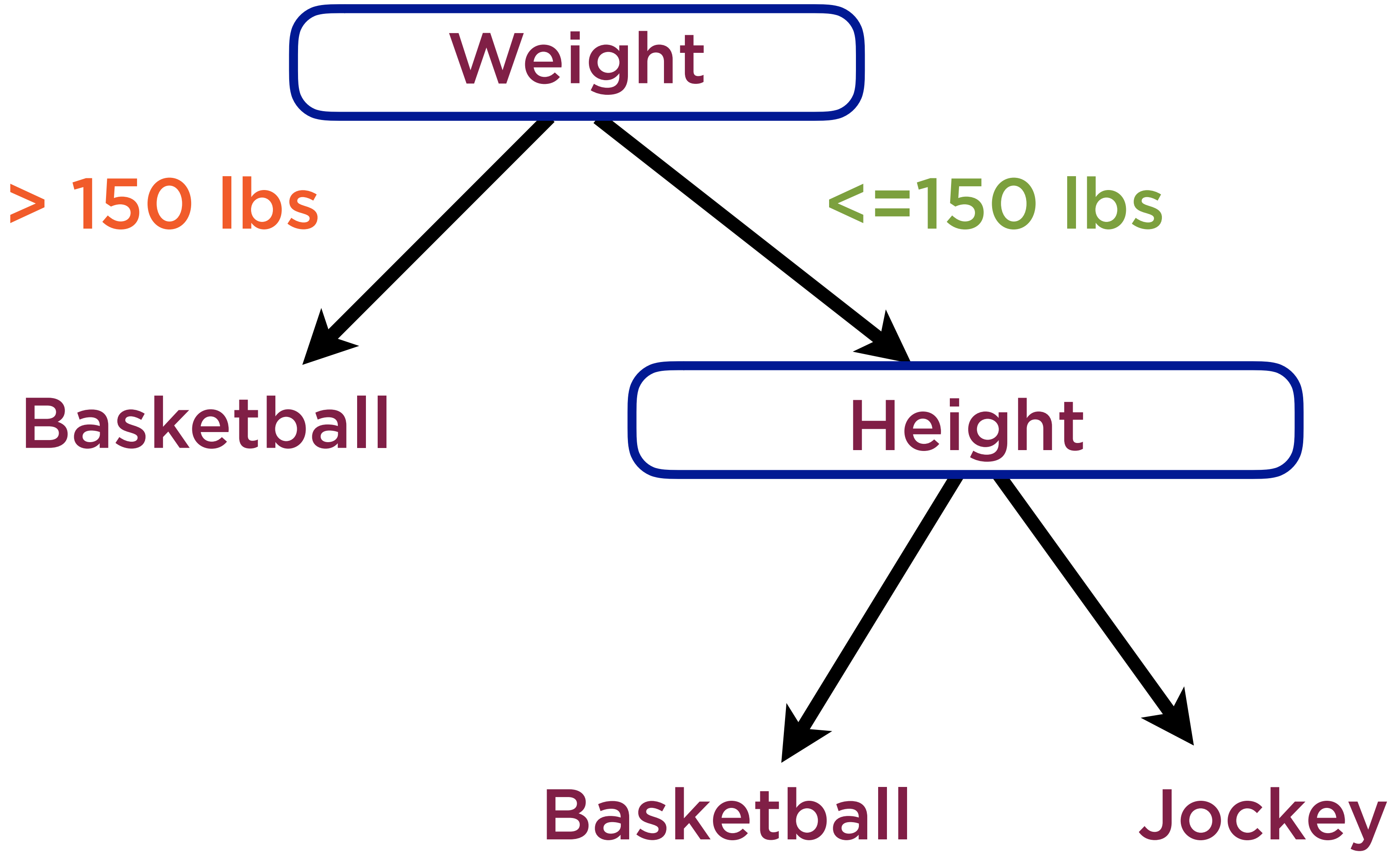
Use rules to make predictions



Decision Tree

“CART”

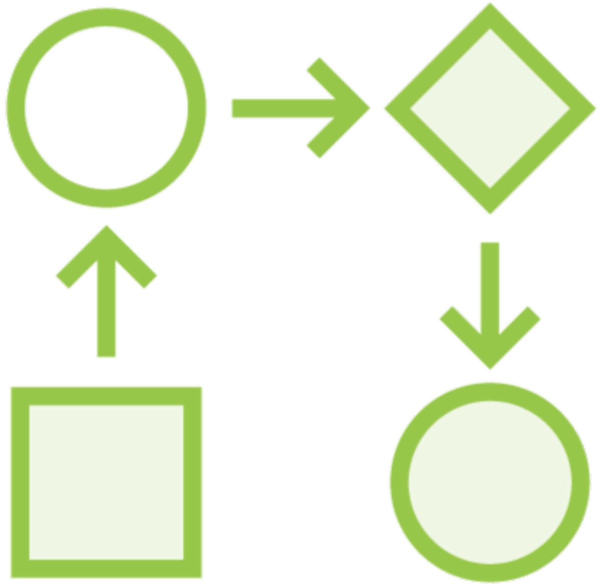
<u>C</u>lassification <u>A</u>nd <u>R</u>egression <u>T</u>ree



Deep Learning Models

Fully-connected, dense neural networks

Convolutional neural networks



Recurrent neural networks

Dense Neural Networks

Work well with numeric features

Traditional classification, regression

Layers of interconnected neurons

All neurons in one layer connected to neurons in the previous and next layers

Convolutional Neural Networks

Specialize in working with image data

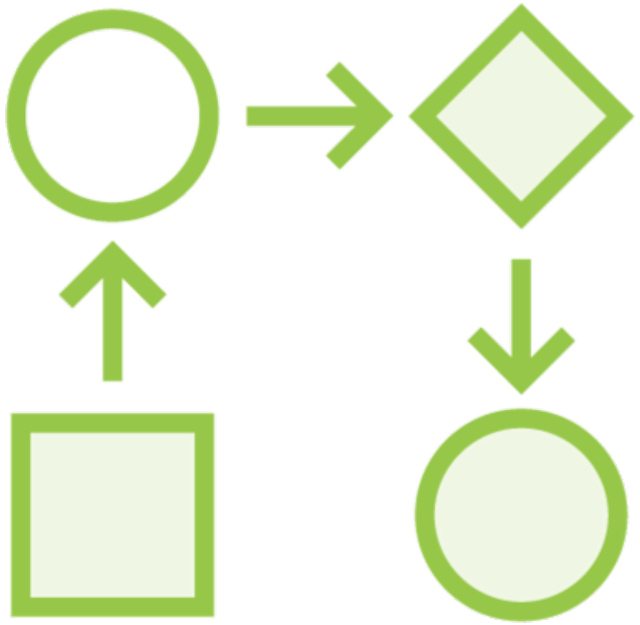
Designed to mimic the visual cortex of the brain

Sparse neural networks

Convolutional layers for feature detection

Pooling layers for subsampling of inputs

Recurrent Neural Networks



Specialize in sequential data such as text or time series data

Neurons have “memory” or state

Neural network layers represent instances in time

Summary

Data-driven decisions and actions

Rule-based approaches to learning

Learning dynamically from changing data using machine learning

Feature extraction from unstructured data using deep learning

Traditional machine learning vs. deep learning