Solving Sentiment Analysis with an ML-based Approach

Vitthal Srinivasan
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

Understand the intuition behind ML-based approaches

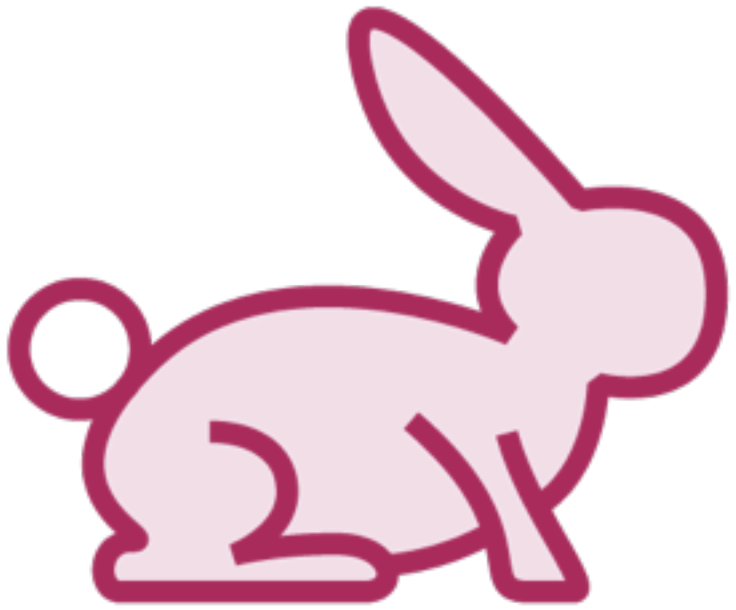
Apply Naive Bayes classification methods to sentiment analysis

Superficially introduce SVM classifiers

Represent documents with different feature vector types

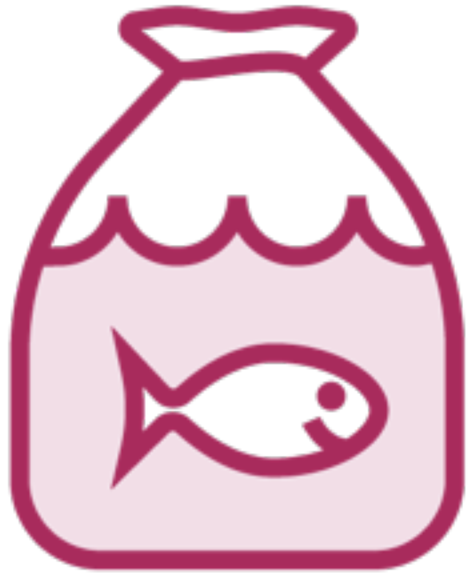
Exploring ML-based Approaches

Whales: Fish or Mammals?



Mammals

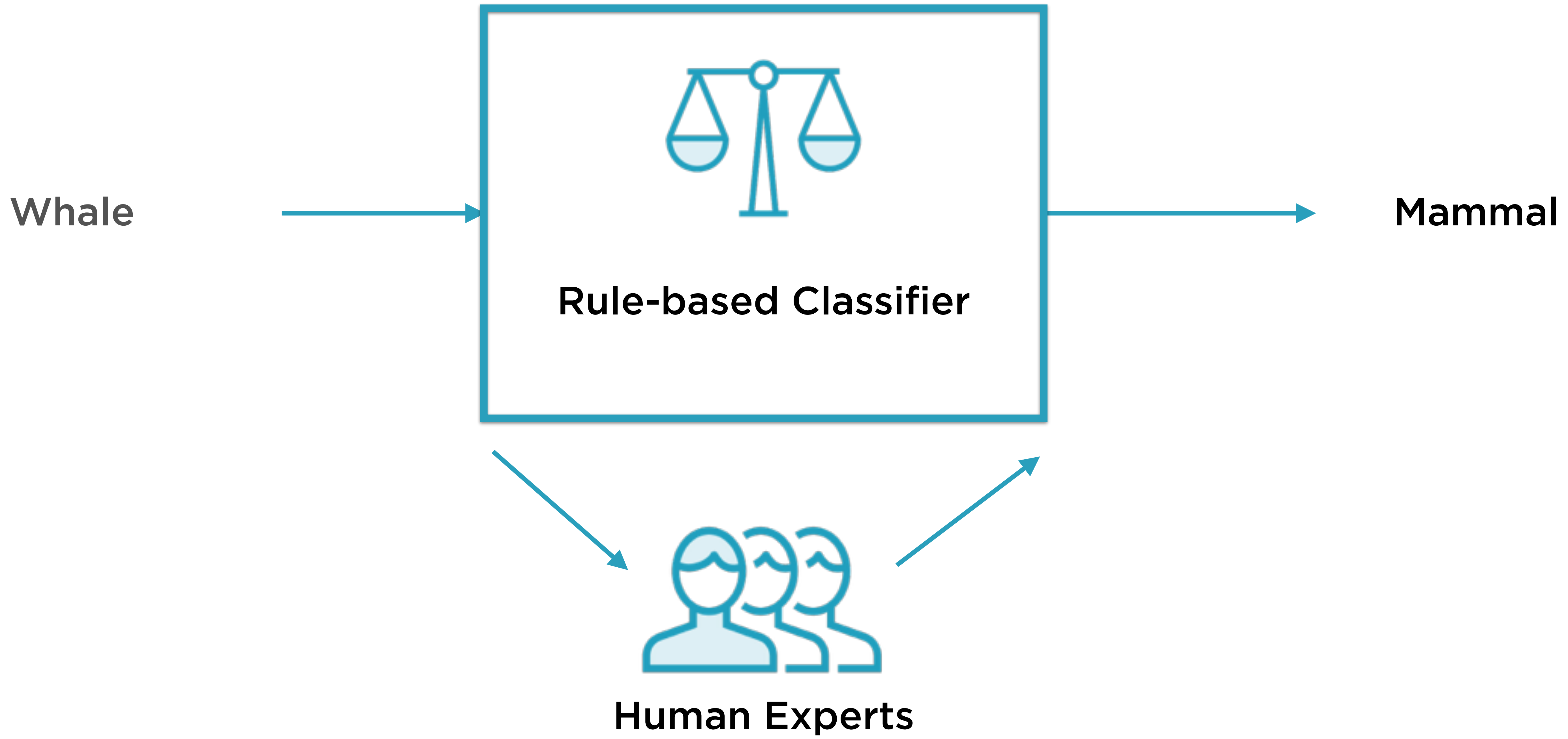
Members of the infraorder *Cetacea*



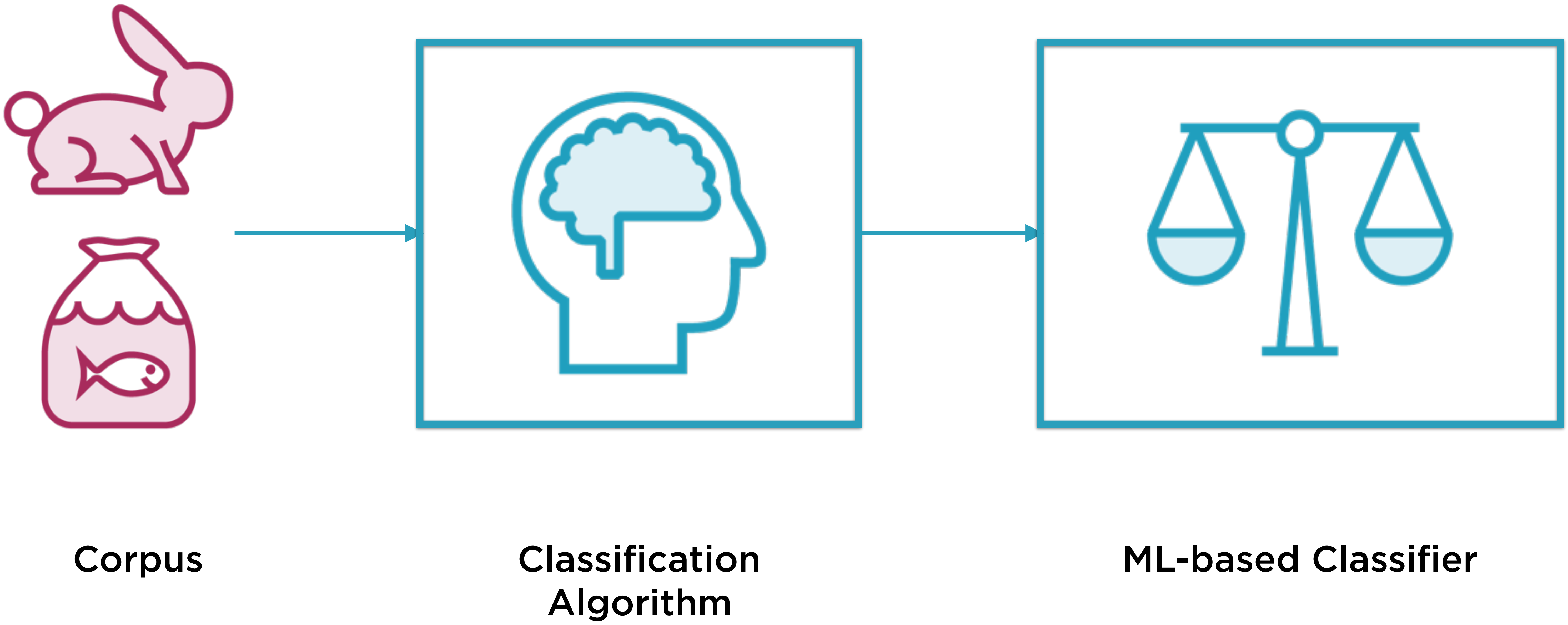
Fish

Look like fish, swim like fish, move with fish

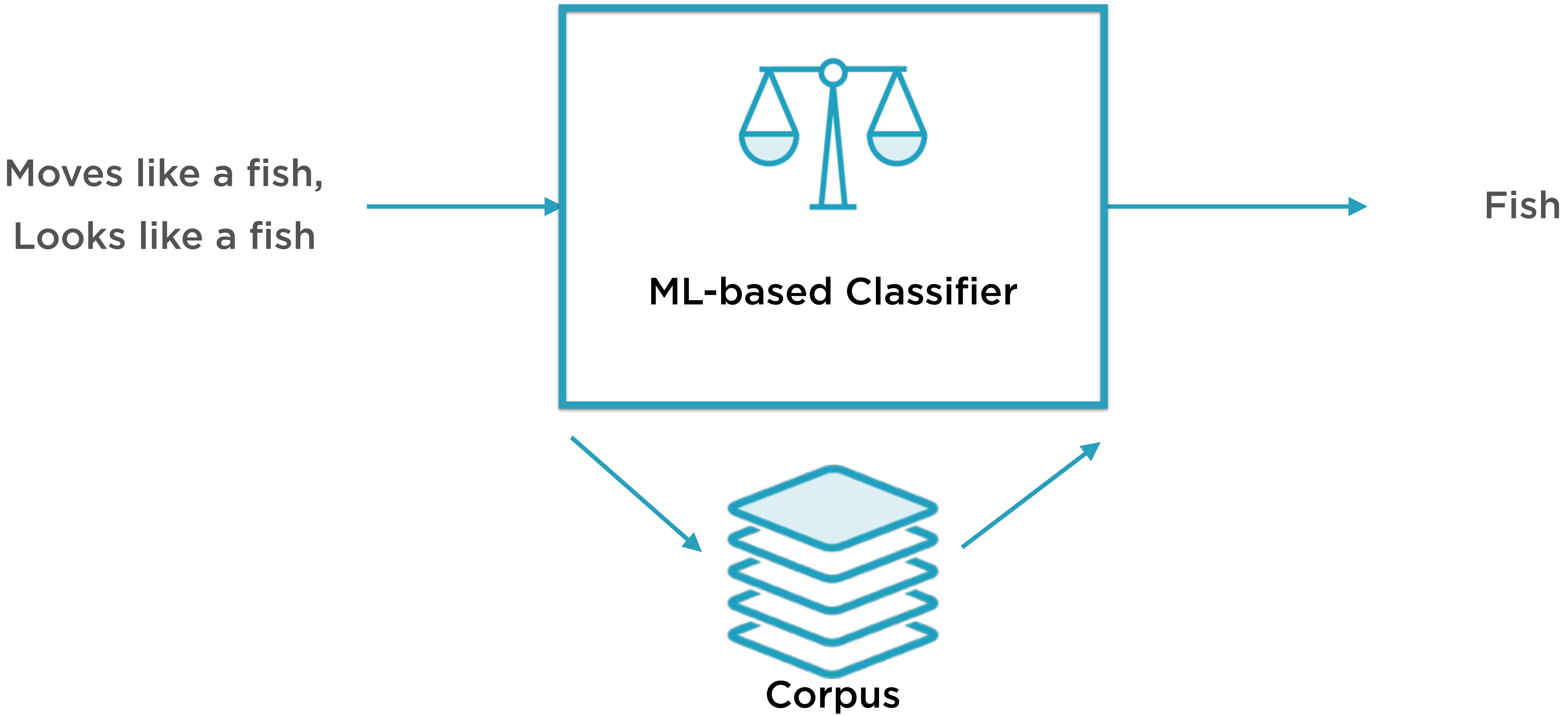
Rule-based Binary Classifier



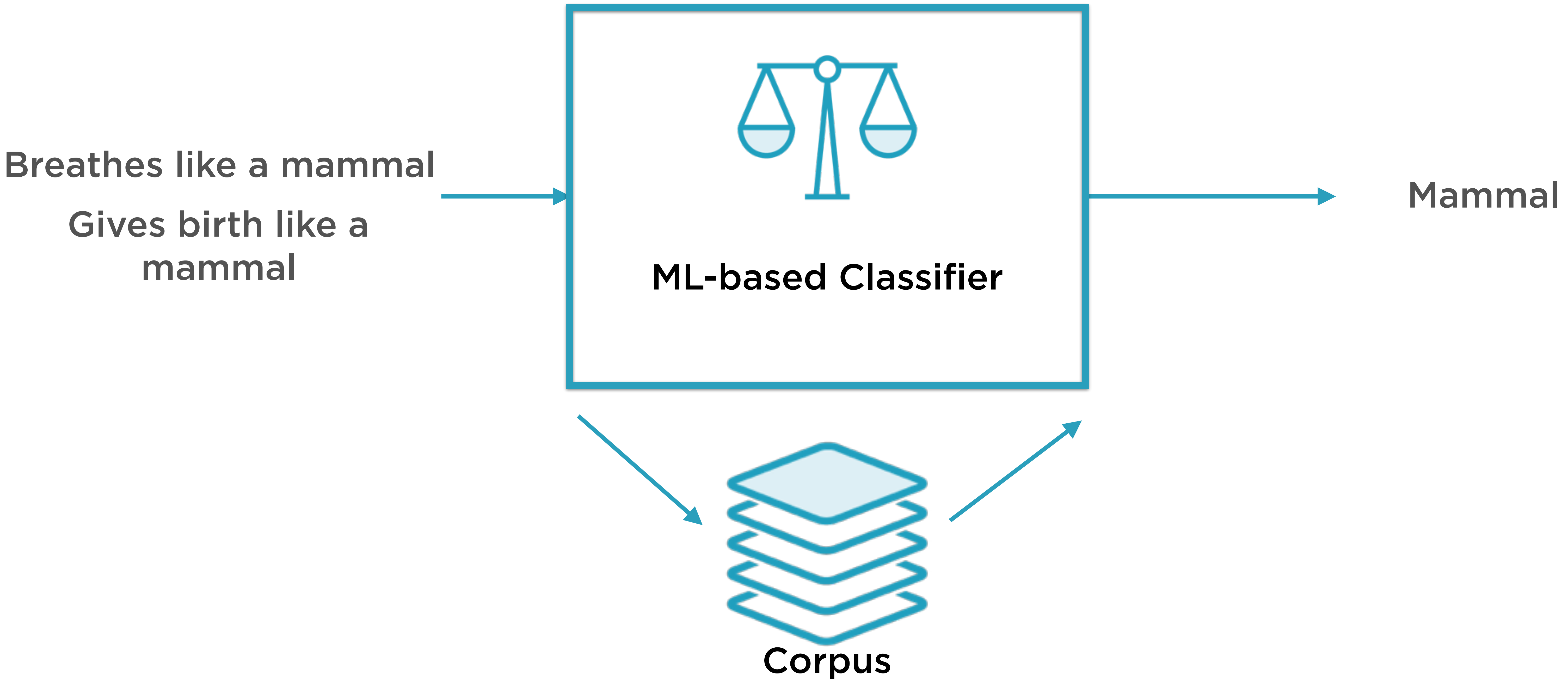
ML-based Binary Classifier



ML-based Binary Classifier

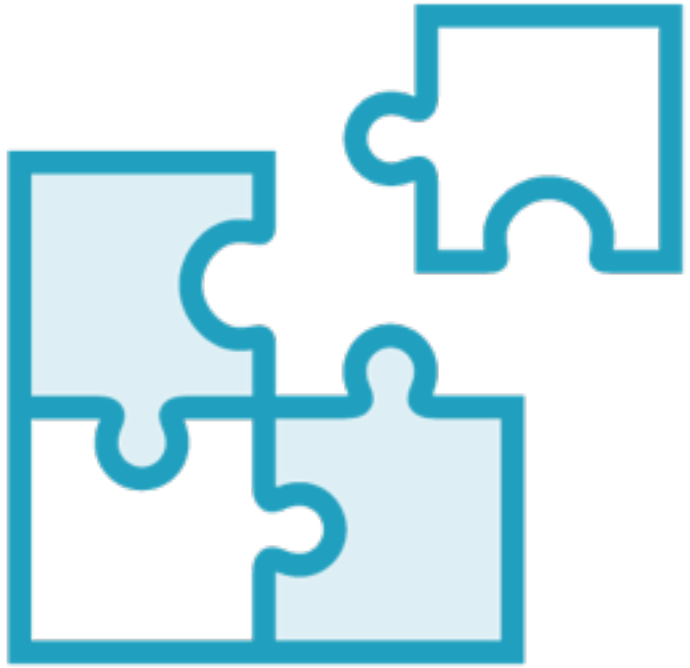


ML-based Binary Classifier



Rule-based or ML-based?

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



Context

Different meanings in different contexts

Rule-based analysers often struggle with context

ML-based systems tend to do better if 'trained' with the right data

The Intuition Behind Bayes Theorem

Swoosh as a Binary Classification Problem

Runner

Police Officer

Classify a person who jogs past you on the street

A Priori Probabilities

Items	P(Occurence)
Runners	9
Police officers	1
Total	10

Observation 1: Today is the city marathon, more runners than police officers out on the streets

A Priori Probabilities

P(Runner) = 9/10

P(Police Officer) = 1/10

These are ***a priori* probabilities:** before anything specific about the person is known

Conditional Probabilities

Handcuffs

Walkie-Talkie

Running Shoes

Observation 2: Specific items appear more often with one category than with the other

Conditional Probabilities

Item	Occurrences with Police Officers	Occurrences with Runners
Handcuffs	6	0
Running Shoes	2	8
Gun	9	0
Badge	8	0
Walkie-Talkie	8	3

Upon Closer Examination

Handcuffs

Badge

The person that zipped past carried these two items

Applying Bayes Theorem

P(Runner/ Handcuffs,Badge)	=	Probability that a person carrying handcuffs and a badge is a runner

Step 1: Find probability that this person is a runner

Applying Bayes Theorem

P(Police Officer/ Handcuffs,Badge)	=	Probability that a person carrying handcuffs and a badge is a police officer

Step 2: Find probability that this person is a police officer

Applying Bayes Theorem

Compare

P(Police Officer/
Handcuffs,Badge)

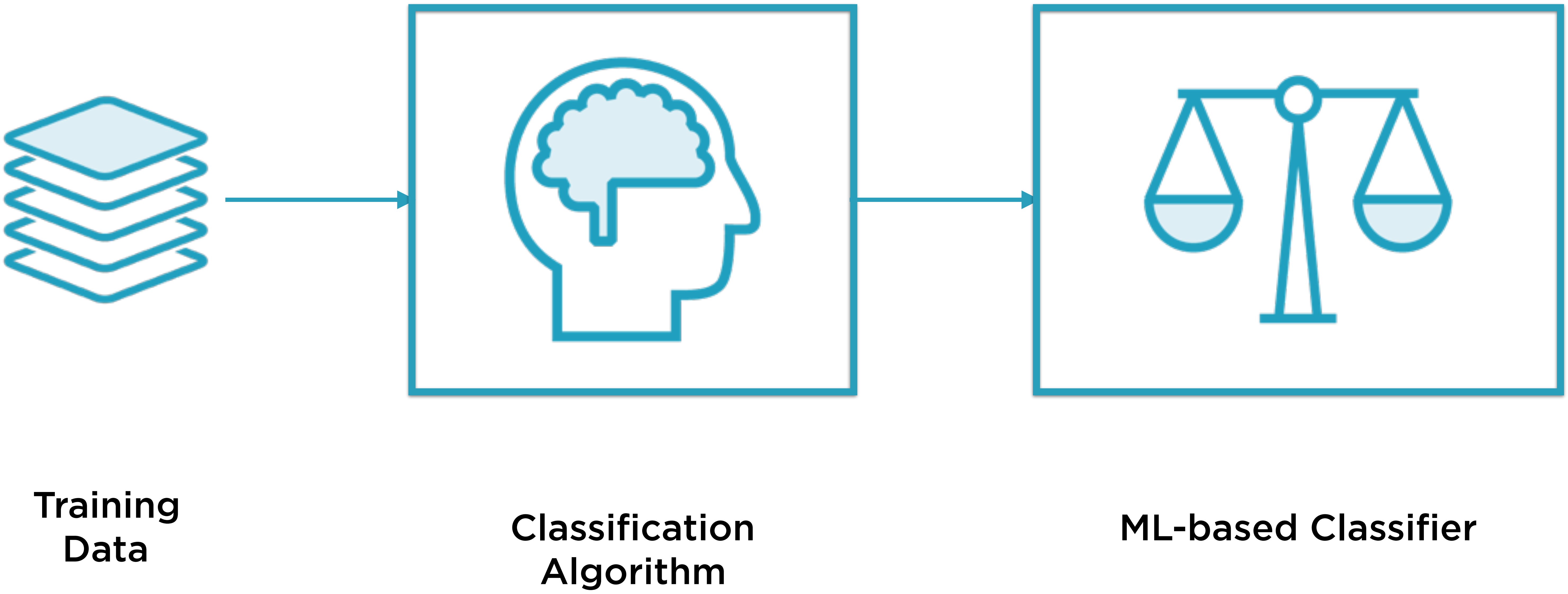
and

P(Runner/
Handcuffs,Badge) =

Step 3: Pick the label with the higher probability

Naive Bayes for Classification Problems

ML-based Binary Classifier



Training Data

Reviews	Labels
Amazing!	Positive
Worst movie ever	Negative
Two thumbs up	Positive
Part 2 was bad, 3 the worst	Negative
Up there with the greats	Positive

Apply Bayes Theorem to probability information from the training data to classify problem instances

A Priori Probabilities

Reviews	Occurence
Positive	3
Negative	2
Total	5

Observation 1: There are more positive reviews than negative reviews in the training data

A Priori Probabilities

Reviews	P(Occurence)
Positive	3/5
Negative	2/5
Total	1

Observation 1: There are more positive reviews than negative reviews in the training data

A Priori Probabilities

P(Positive) = 3/5

P(Negative) = 2/5

These are *a priori* probabilities: before anything specific about review contents is known

Conditional Probabilities

Reviews
Amazing!
Worst movie ever
Two thumbs up
Part 2 was bad, 3 the worst
Up there with the greats

Labels
Positive
Negative
Positive
Negative
Positive

Observation 2: Specific words occur more in one type of review than in the other

Conditional Probabilities

Reviews	Labels
Amazing!	Positive
Worst movie ever	Negative
Two thumbs up	Positive
Part 2 was bad, 3 the worst	Negative
Up there with the greats	Positive

The word up appears twice in positive reviews, but zero times in negative reviews

Conditional Probabilities

Reviews	Labels
Amazing!	Positive
Worst movie ever	Negative
Two thumbs up	Positive
Part 2 was bad, 3 the worst	Negative
Up there with the greats	Positive

The word worst appears twice in negative reviews, and zero times in positive reviews

Conditional Probabilities

Word	Occurrences in Positive Reviews	Occurrences in Negative Reviews
amazing	1	
worst		2
movie		1
ever		1
two	1	1
thumbs	1	
up	2	
Part		1
was		1
bad		1
3		1
the	1	1
there	1	
with	1	
greats	1	
	9	10

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	

P(text contains "amazing"/label = Positive) = 1/9

P(text contains "amazing"/label = Negative) = 0

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

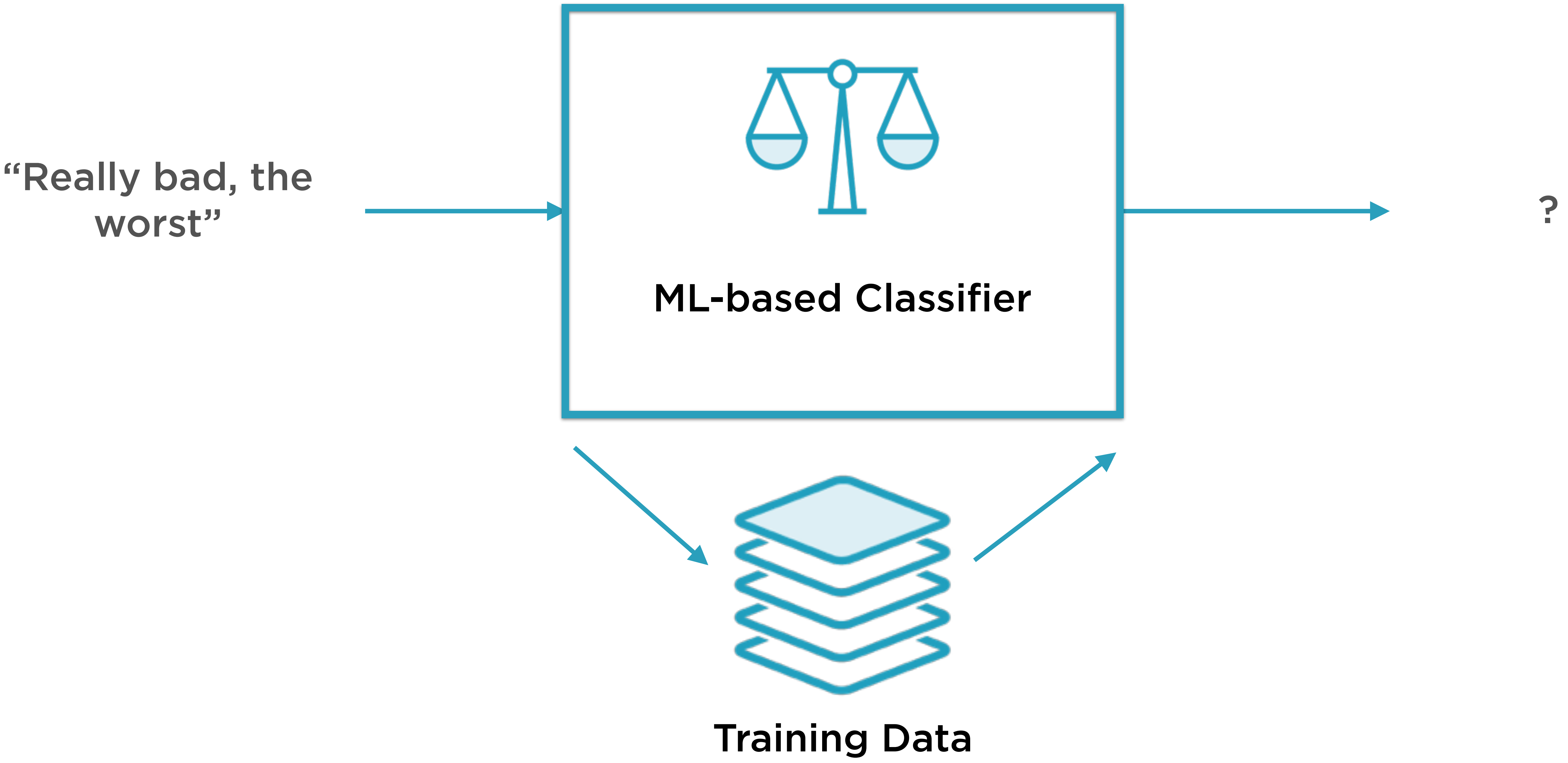
Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
worst		2/10

P(text contains "worst"/label = Positive) = 0

P(text contains "worst"/label = Negative) = 2/10

Classifying a New Problem Instance



Classifying a New Problem Instance

Reviews
Amazing!
Worst movie ever
Two thumbs up
Part 2 was bad, 3 the worst
Up there with the greats

Labels
Positive
Negative
Positive
Negative
Positive

"Really bad, the worst"

Given the words in this review, call them *t*, is the review likely to be positive or negative?

Applying Bayes Theorem

P(Positive/t) = P(label = Positive/ text = “Really bad, the worst”)

Step 1: Find probability that the review is actually positive, given the text of the review (use Bayes Theorem)

Applying Bayes Theorem

P(Negative/t) = P(label = Negative/ text = “Really bad, the worst”)

Step 2: Find probability that the review is actually negative, given the text of the review (use Bayes Theorem)

Applying Bayes Theorem

```
If P(Positive/t) > P(t/Negative/t)
      classify t as Positive
else
      classify t as Negative
```

Step 3: Pick the label with the higher probability

Naive Bayes Classification

P(Positive/t) = P(label = Positive/ text = "Really bad, the worst")

P(Negative/t) = P(label = Negative/ text = "Really bad, the worst")

If P(Positive/t) > P(t/Negative/t)
    classify t as Positive
else
    classify t as Negative

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times P(Positive)}{P(t/Positive) \times P(Positive) + P(t/Negative) \times P(Negative)}$$

Step 1: Find probability that the review is actually positive, given the text of the review (use Bayes Theorem)

Applying Bayes Theorem

$$P(Negative/t) = \frac{P(t/Negative) \times P(Negative)}{P(t/Positive) \times P(Positive) + P(t/Negative) \times P(Negative)}$$

Step 2: Find probability that the review is actually negative, given the text of the review (use Bayes Theorem)

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times P(Positive)}{P(t/Positive) \times P(Positive) + P(t/Negative) \times P(Negative)}$$

$$P(Negative/t) = \frac{P(t/Negative) \times P(Negative)}{P(t/Positive) \times P(Positive) + P(t/Negative) \times P(Negative)}$$

Applying Bayes Theorem

$$P(\text{Positive}/t) = \frac{P(t/\text{Positive}) \times P(\text{Positive})}{P(t/\text{Positive}) \times P(\text{Positive}) + P(t/\text{Negative}) \times P(\text{Negative})}$$

$$P(\text{Negative}/t) = \frac{P(t/\text{Negative}) \times P(\text{Negative})}{P(t/\text{Positive}) \times P(\text{Positive}) + P(t/\text{Negative}) \times P(\text{Negative})}$$

Applying Bayes Theorem

$$P(\text{Positive}/t) = \frac{P(t/\text{Positive}) \times P(\text{Positive})}{\ldots}$$

$$P(\text{Negative}/t) = \frac{P(t/\text{Negative}) \times P(\text{Negative})}{\ldots}$$

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times P(Positive)}{...}$$

$$P(Negative/t) = \frac{P(t/Negative) \times P(Negative)}{...}$$

A Priori Probabilities

P(Positive) = 3/5

Before we know anything about review contents

P(Negative) = 2/5

Before we know anything about review contents

Observation 1: There are more positive reviews than negative reviews in the training data

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times P(Positive)}{\ldots}$$

$$P(Negative/t) = \frac{P(t/Negative) \times P(Negative)}{\ldots}$$

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times 3/5}{\ldots}$$

$$P(Negative/t) = \frac{P(t/Negative) \times 2/5}{\ldots}$$

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times 3/5}{\ldots}$$

$$P(Negative/t) = \frac{P(t/Negative) \times 2/5}{\ldots}$$

Applying Bayes Theorem

P(t/Positive) = P(text = “Really bad, the worst” /label = Positive)

= P(text contains “Really”/label = Positive) AND

P(text contains “bad”/label = Positive) AND

P(text contains “the”/label = Positive) AND

P(text contains “worst”/label = Positive)

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= P(text contains "Really"/label = Positive) AND

P(text contains "bad"/label = Positive) AND

P(text contains "the"/label = Positive) AND

P(text contains "worst"/label = Positive)

Naive Bayes makes naive (strong) assumptions about independence of features

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= P(text contains "Really"/label = Positive) AND

P(text contains "bad"/label = Positive) AND

P(text contains "the"/label = Positive) AND

P(text contains "worst"/label = Positive)

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= P(text contains "Really"/label = Positive) x

P(text contains "bad"/label = Positive) x

P(text contains "the"/label = Positive) x

P(text contains "worst"/label = Positive)

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= P(text contains "Really"/label = Positive) x

P(text contains "bad"/label = Positive) x

P(text contains "the"/label = Positive) x

P(text contains "worst"/label = Positive)

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
amazing	1/9	
worst		2/10
movie		1/10
ever		1/10
two	1/9	1/10
thumbs	1/9	
up	2/9	
Part		1/10
was		1/10
bad		1/10
3		1/10
the	1/9	1/10
there	1/9	
with	1/9	
greats	1/9	

Conditional Probabilities

Word	P(Occurrences in Positive Reviews)	P(Occurrences in Negative Reviews)
worst		2/10

P(text contains “worst”/label = Positive) = 0

P(text contains “worst”/label = Negative) = 2/10

Applying Bayes Theorem

P(t/Negative) = P(text = "Really bad, the worst" /label = Negative)

= ~~P(text contains "Really"/label = Negative) x~~

1/10 x

1/10 x

2/10

= **2/1000**

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= P(text contains "Really"/label = Positive) x

P(text contains "bad"/label = Positive) x

P(text contains "the"/label = Positive) x

P(text contains "worst"/label = Positive)

Applying Bayes Theorem

P(t/Positive) = P(text = "Really bad, the worst" /label = Positive)

= ~~P(text contains "Really"/label = Positive) x~~

0 x

1/10 x

0

= **0**

Applying Bayes Theorem

$$P(Positive/t) = \frac{P(t/Positive) \times 3/5}{\ldots}$$

$$P(Negative/t) = \frac{P(t/Negative) \times 2/5}{\ldots}$$

Applying Bayes Theorem

$$P(\text{Positive}/t) = \frac{0 \times 3/5}{\ldots}$$

$$P(\text{Negative}/t) = \frac{2/1000 \times 2/5}{\ldots}$$

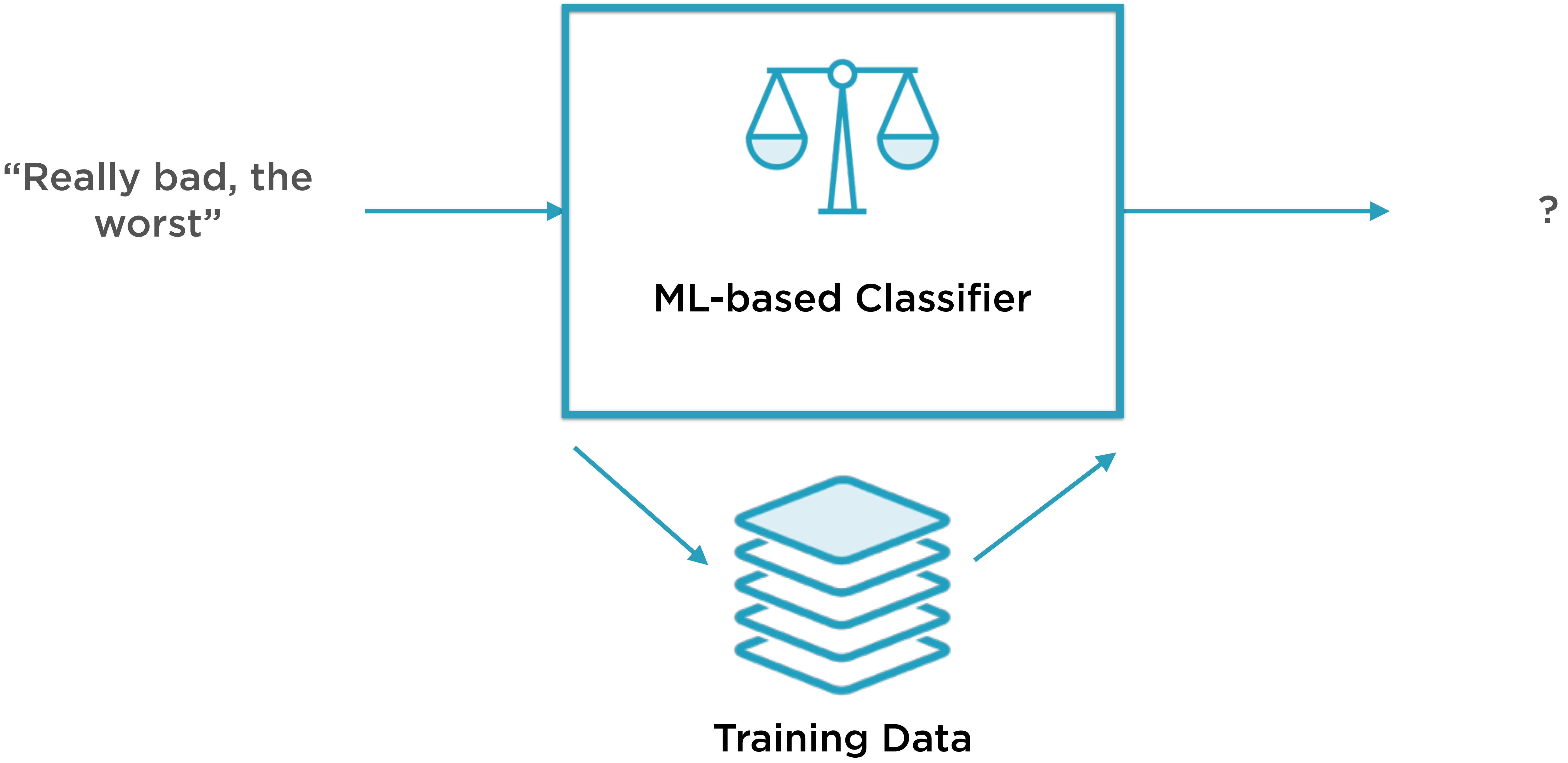
Applying Bayes Theorem

P(Positive/t) = P(label = Positive/ text = "Really bad, the worst")

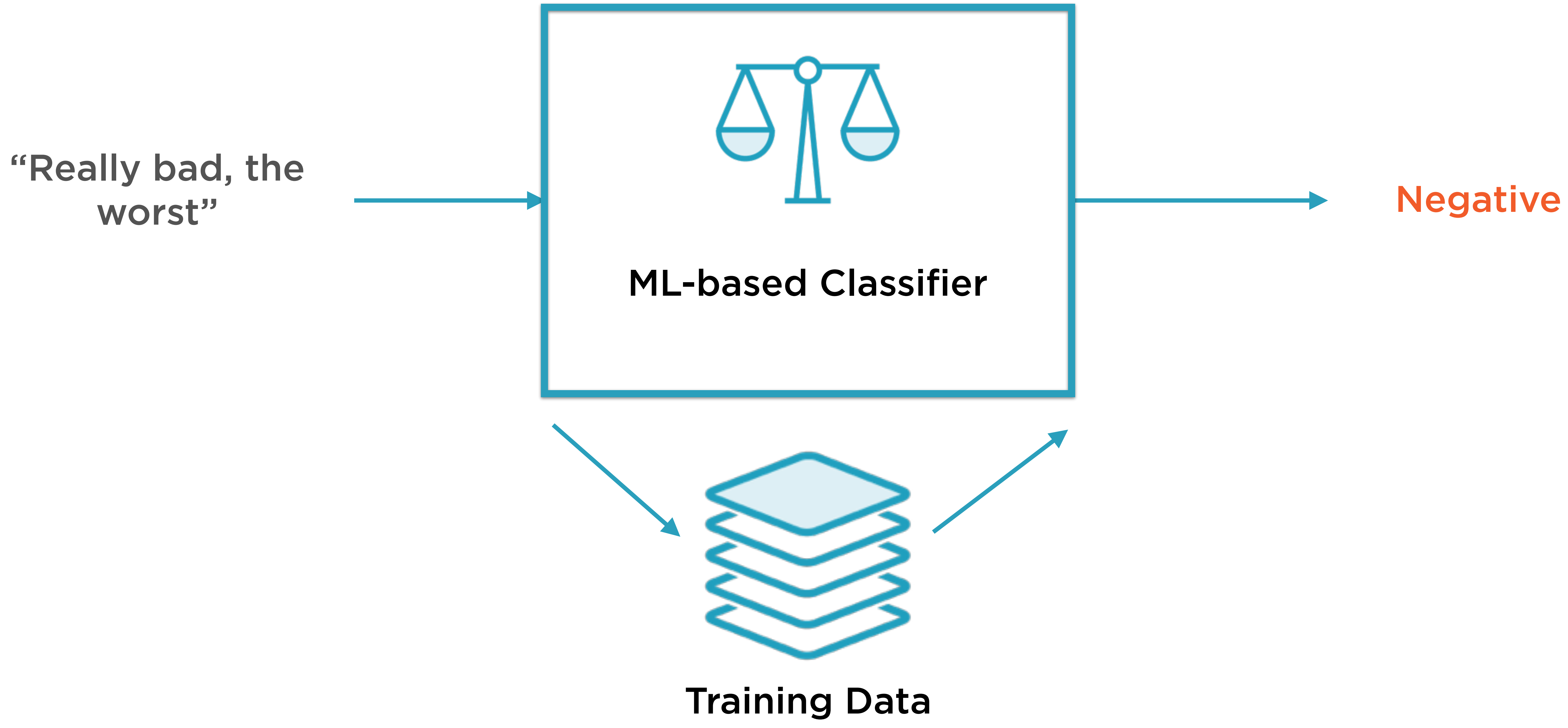
P(Negative/t) = P(label = Negative/ text = "Really bad, the worst")

If P(Positive/t) > P(t/Negative/t)
    classify t as Positive
else
    classify t as Negative

Classifying a New Problem Instance



Classifying a New Problem Instance



Support Vector Machines for Classification Problems

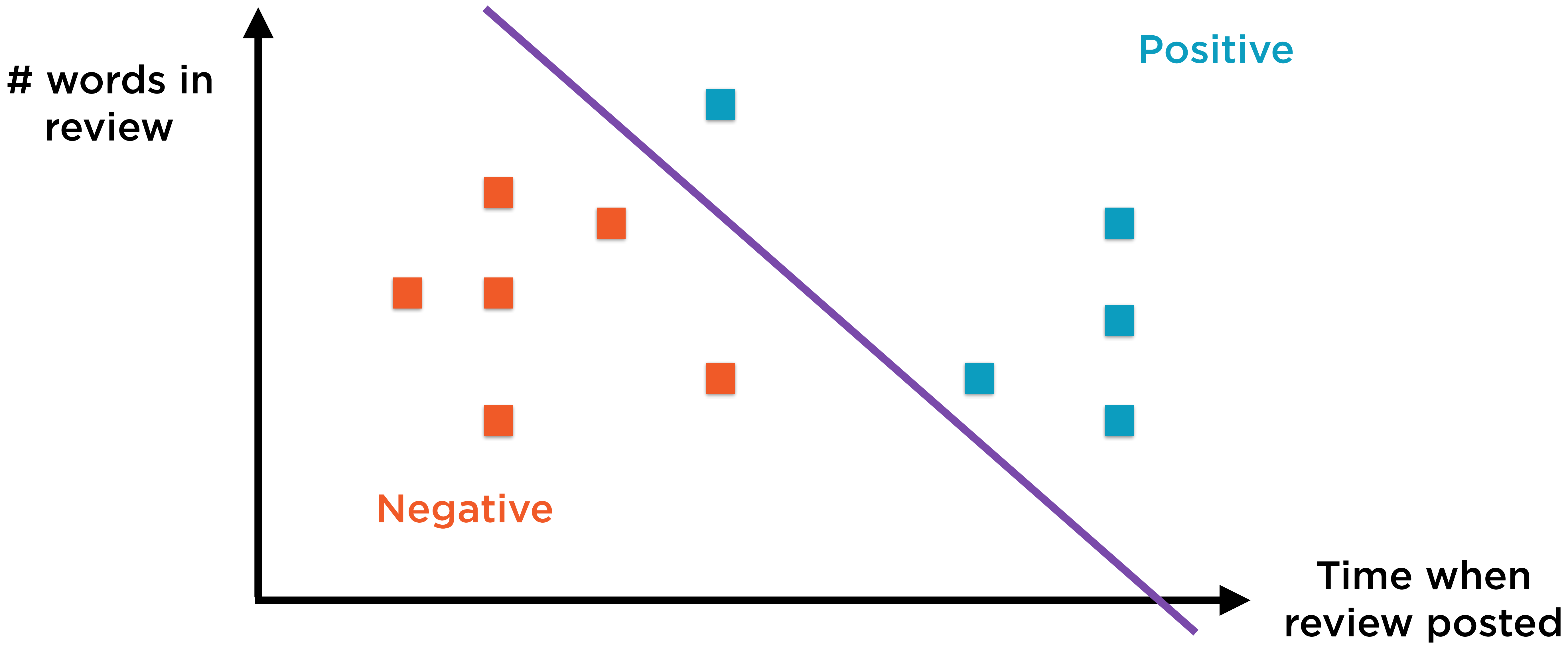
Data in One Dimension

Unidimensional data points can be represented using a line, such as a number line

Data in One Dimension

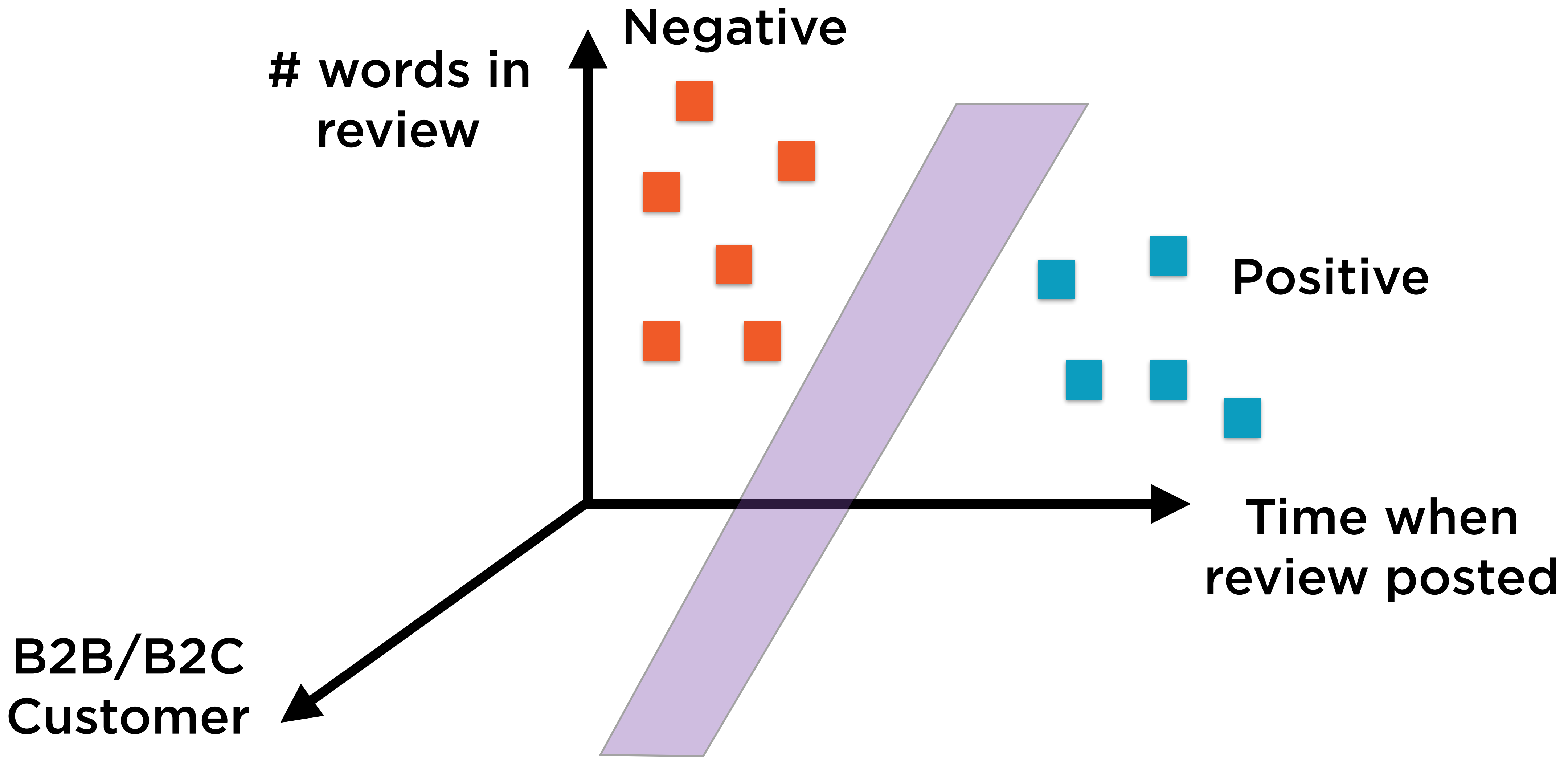
Unidimensional can also be separated, or classified, using a point

Data in Two Dimensions



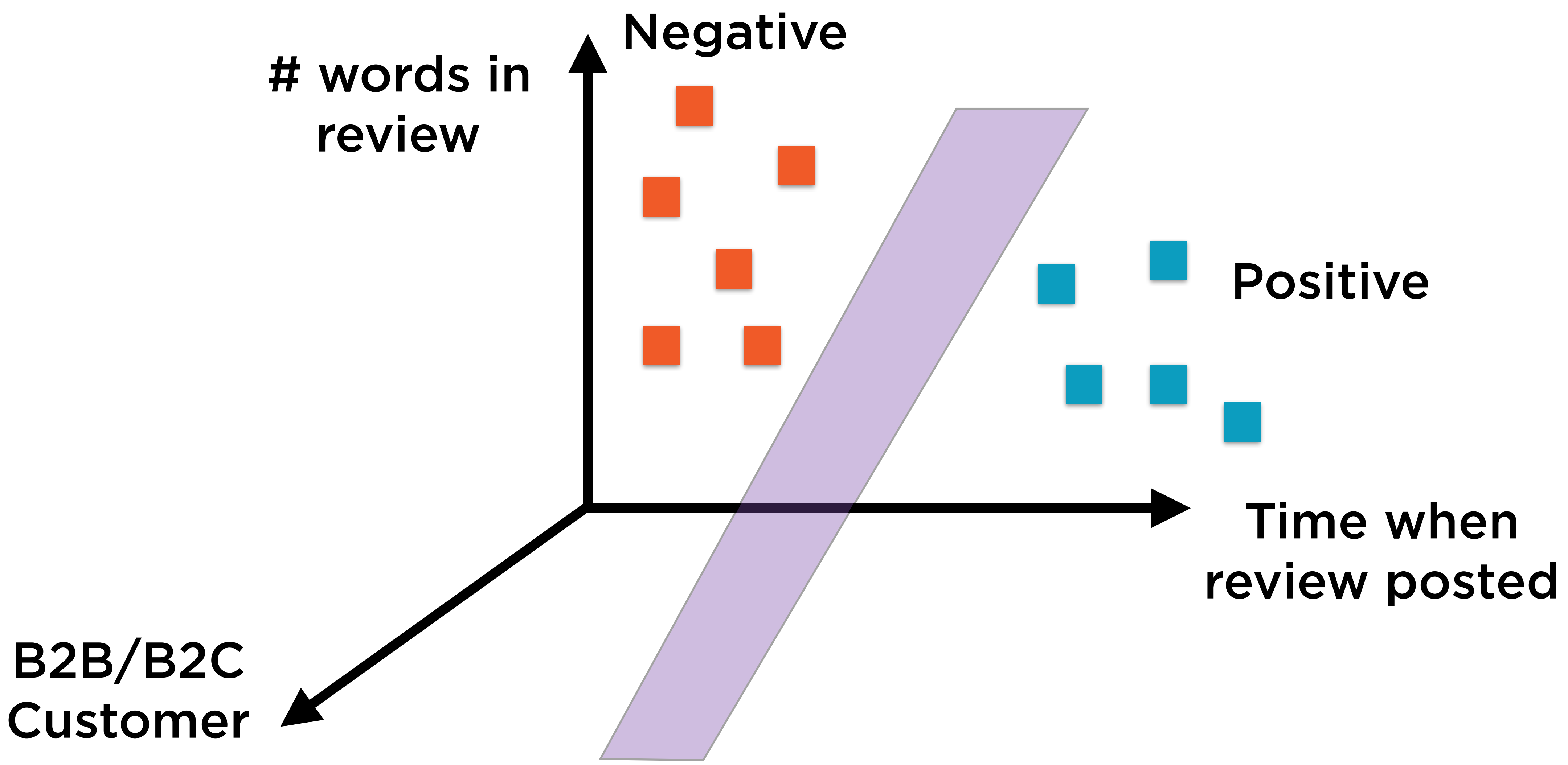
Bidimensional data points can be represented using a plane, and classified using a line

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

SVM

SVM classifiers are incredibly hard to build, but fairly easy to use

Kernel functions are key to building SVM classifiers

Feature vectors are key to using SVM classifiers

SVM Is Cool, Naive Bayes Is Robust

SVM is much more complicated

SVM is harder to understand and to use

Naive Bayes often works as well

Even with minimal tuning or tweaking

SVM needs a lot of tuning

Parameter tuning makes all the difference in SVM

Summary

ML-based approaches often involve a crucial training step

Naive Bayes classifiers are simple, robust and perform very well

Other ML-based methods include SVM classifiers

Different methods rely on different feature representations