Understanding and Implementing Classification Models

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

Overview

Binary vs. multiclass classification

Logistic regression intuition

Other classification algorithms

Support vector classification

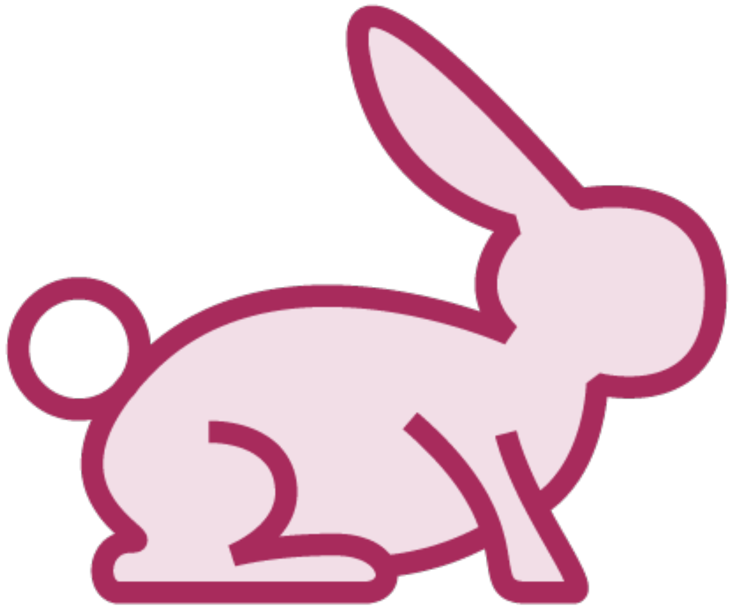
Nearest-neighbors classification

Decision trees for classification

Naive Bayes classification

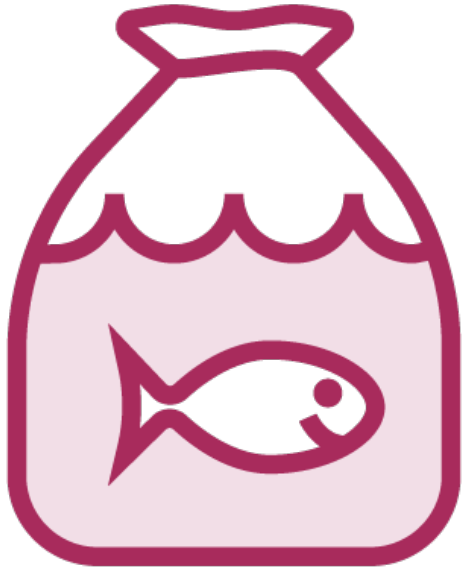
Types of Classification

Whales: Fish or Mammals?



Mammals

Members of the infraorder *Cetacea*



Fish

Look like fish, swim like fish, move with fish

Types of Classification Tasks

Binary

"Yes/No", "True/False", "Up/Down"

Output is binary categorical variable

Multiclass

Digit classification

Output variable takes 1 of N (>2) values

Multilabel

("True", "Female"), ("False", "Female")

Output is tuple of multiple binary variables (not disjoint)

Multioutput

("Sunday", "January")

Multiclass + multilabel

Multilabel

Some algorithms are inherently multilabel

- Naive Bayes

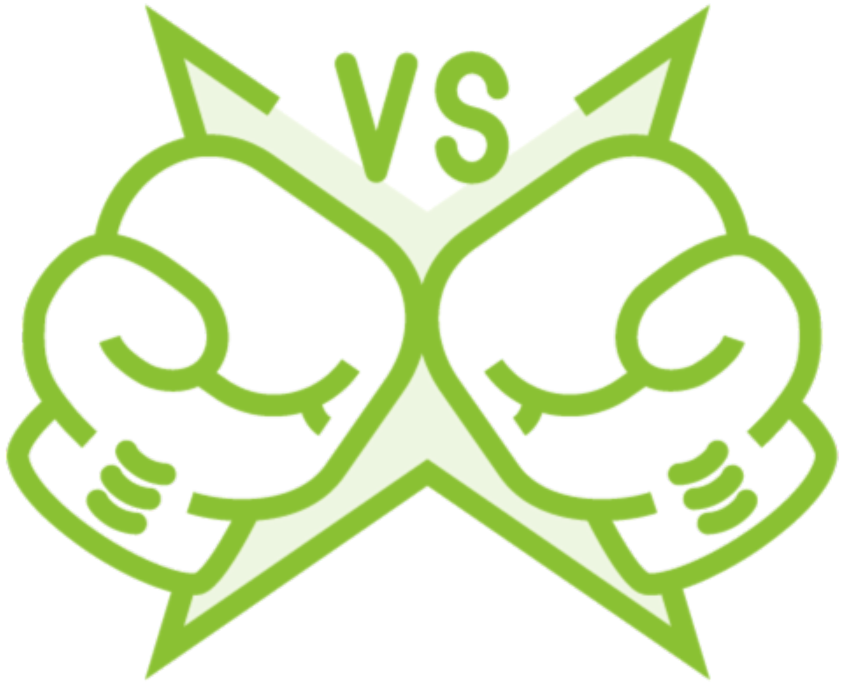
Multiclass

Many classification algorithms are inherently binary

- Logistic regression
- Support Vector Machines

Inherently binary classifiers can be generalized for multiclass classification

One vs. All



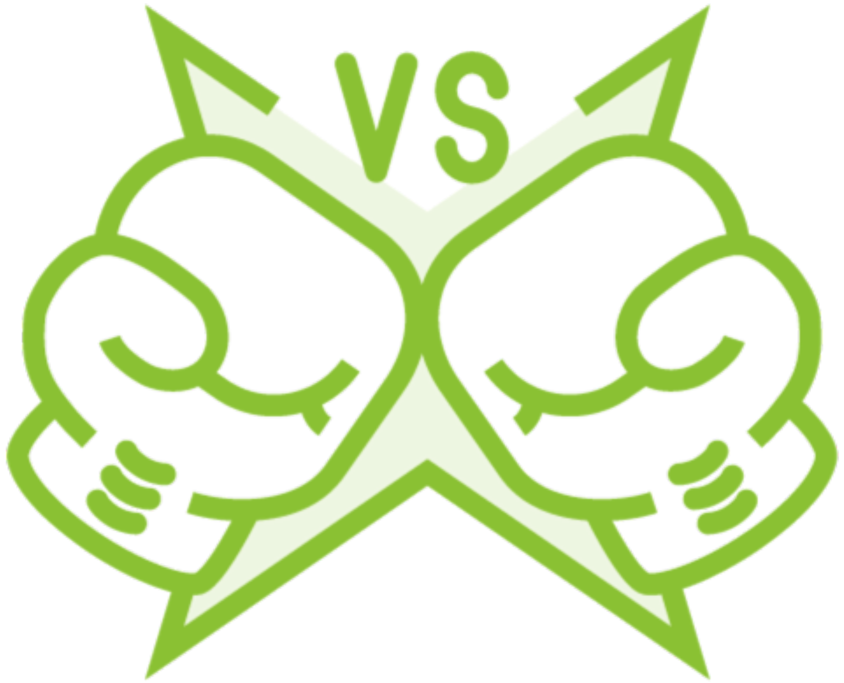
One-versus-all

Classifying digits 0-9

Train 10 binary classifiers

- 0-detector, 1-detector...
- Predicted label = output of detector with highest score

One vs. One



One-versus-one

Train 45 binary classifiers

- One detector for each pair of digits
- For N labels, need N(N-1)/2 classifiers
- Predicted label = output of digit that wins most duels

Logistic Regression: Intuition

Two Approaches to Deadlines

Start 5 minutes before deadline

Good luck with that

Start 1 year before deadline

Maybe overkill

Neither approach is optimal

Starting a Year in Advance

Probability of meeting the deadline

100%

Probability of getting other important work done

0%

Starting Five Minutes in Advance

Probability of meeting the deadline

0%

Probability of getting other important work done

100%

The Goldilocks Solution

Work fast

Start very late and hope for the best

Work smart

Start as late as possible to be sure to make it

Work hard

Start very early and do little else

As usual, the middle path is best

Working Smart

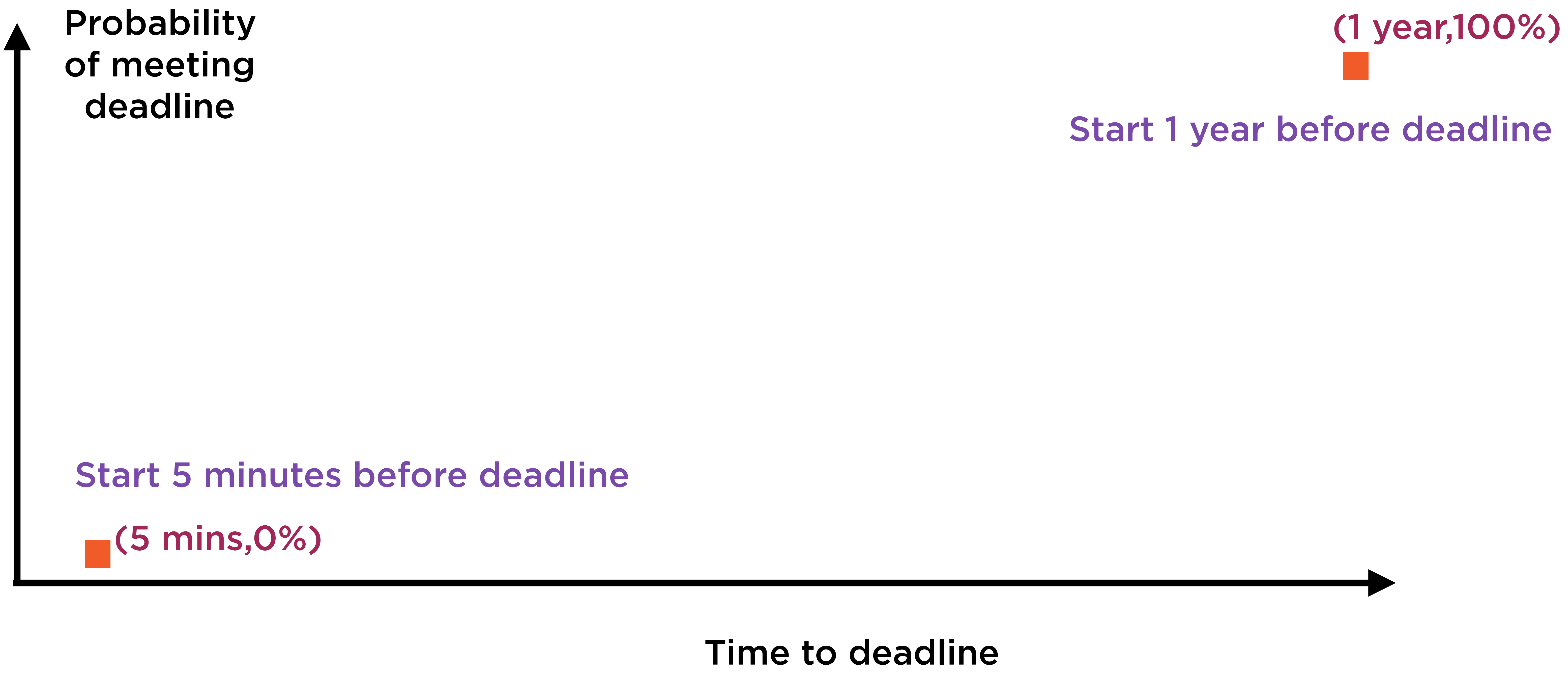
Probability of meeting the deadline

95%

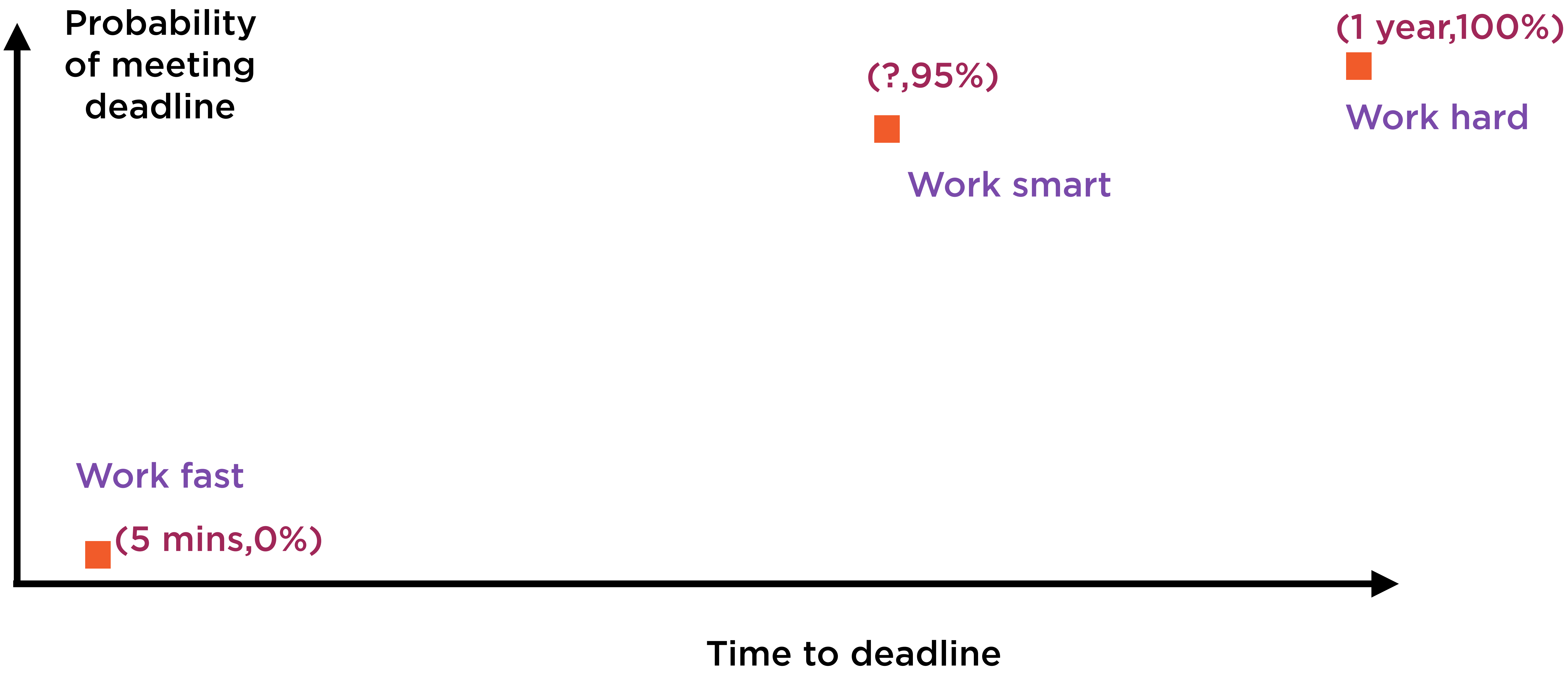
Probability of getting other important work done

95%

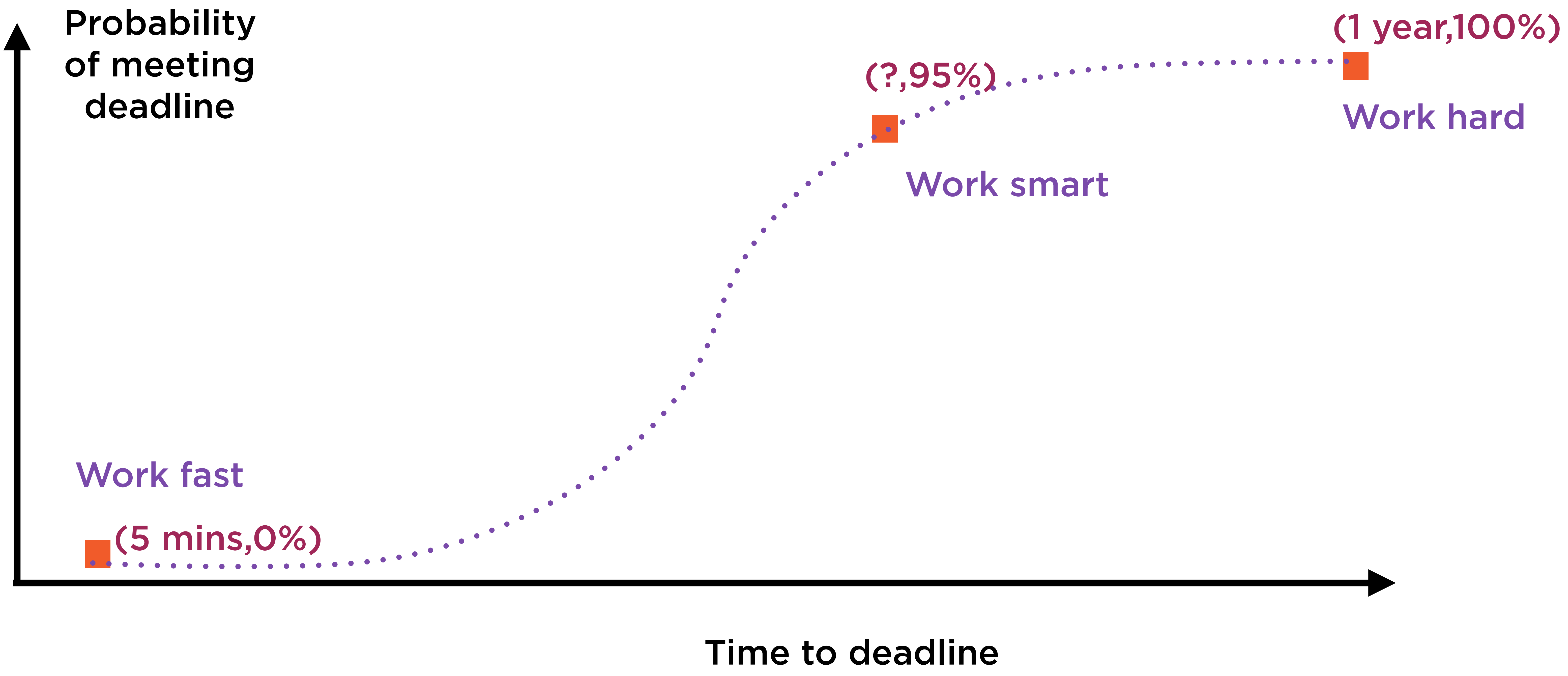
Working Hard, Fast, Smart



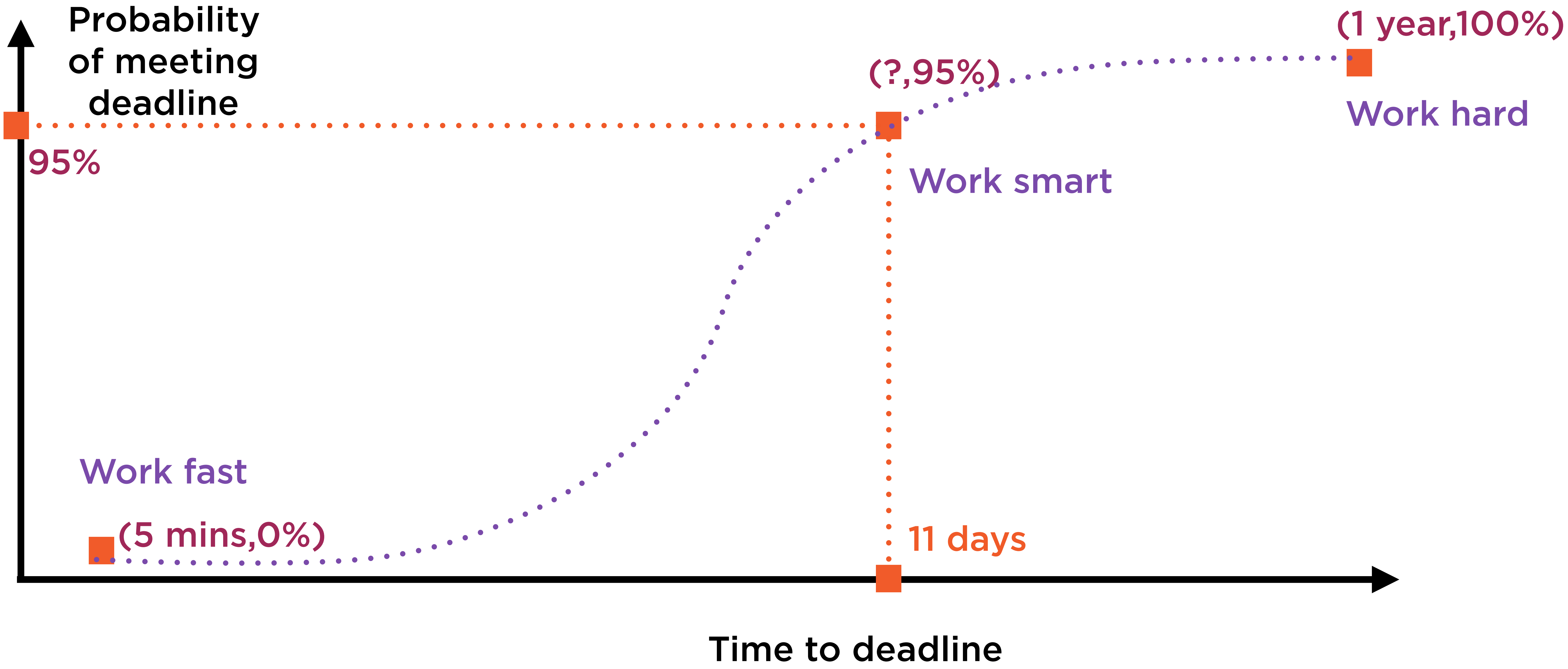
Working Hard, Fast, Smart



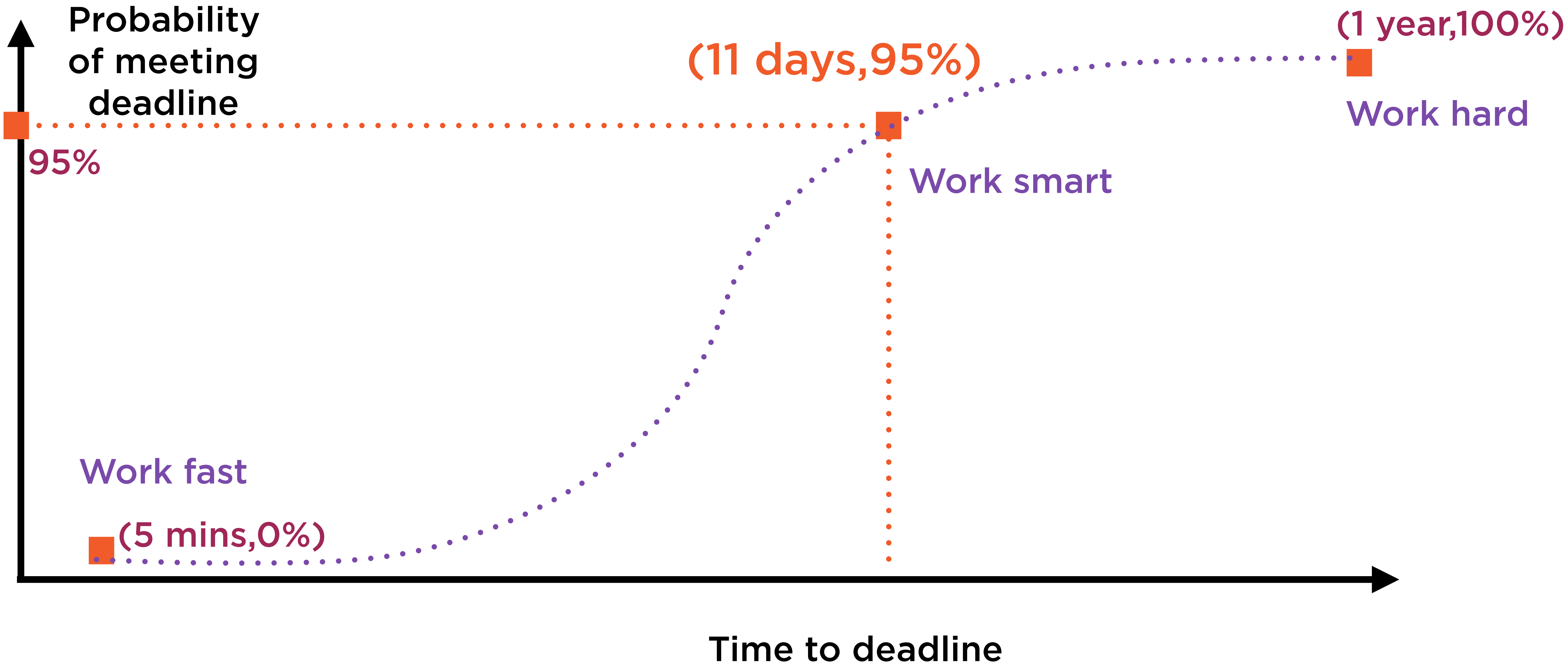
Working Hard, Fast, Smart



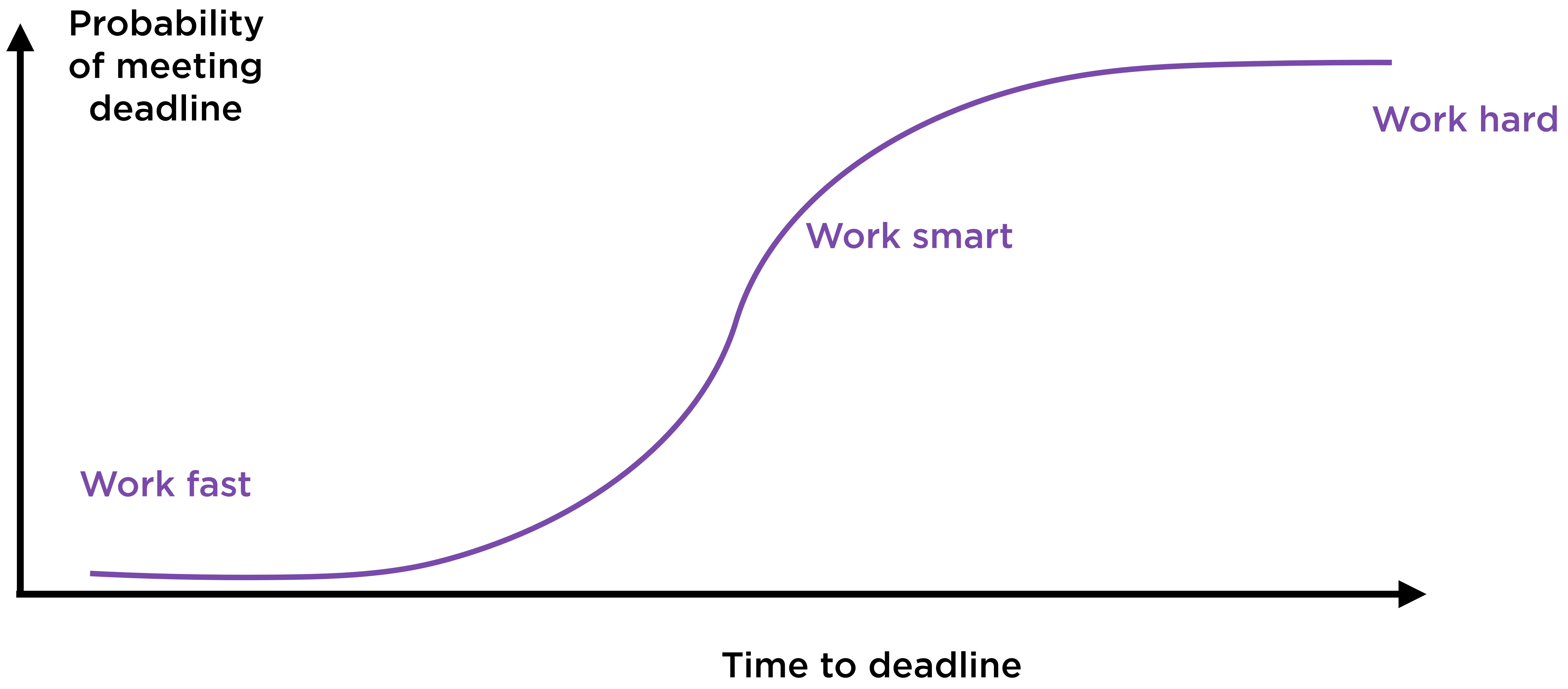
Working Hard, Fast, Smart



Working Hard, Fast, Smart

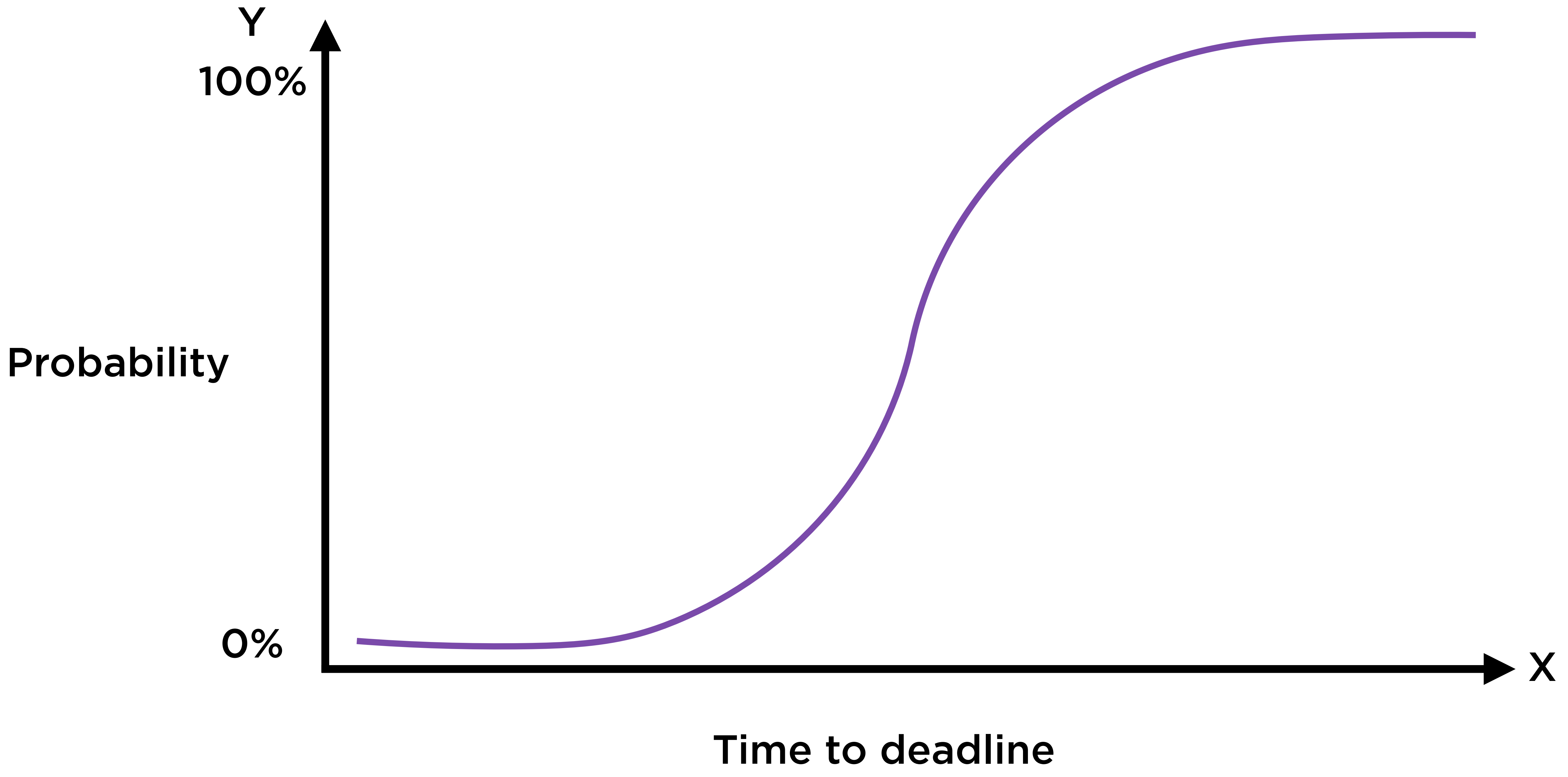


Working Hard, Fast, Smart

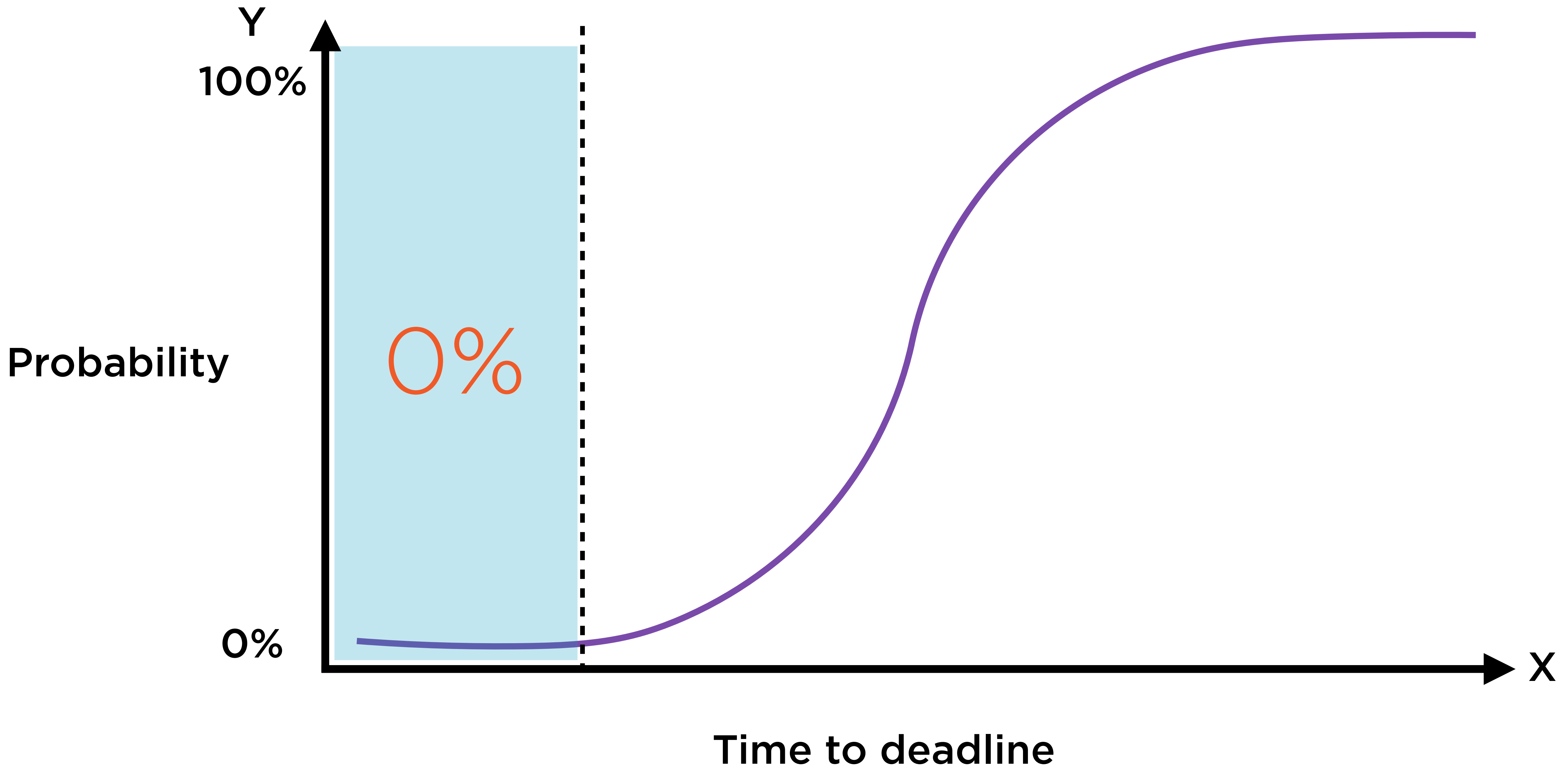


Logistic Regression helps find how probabilities are changed by actions

Working Smart with Logistic Regression

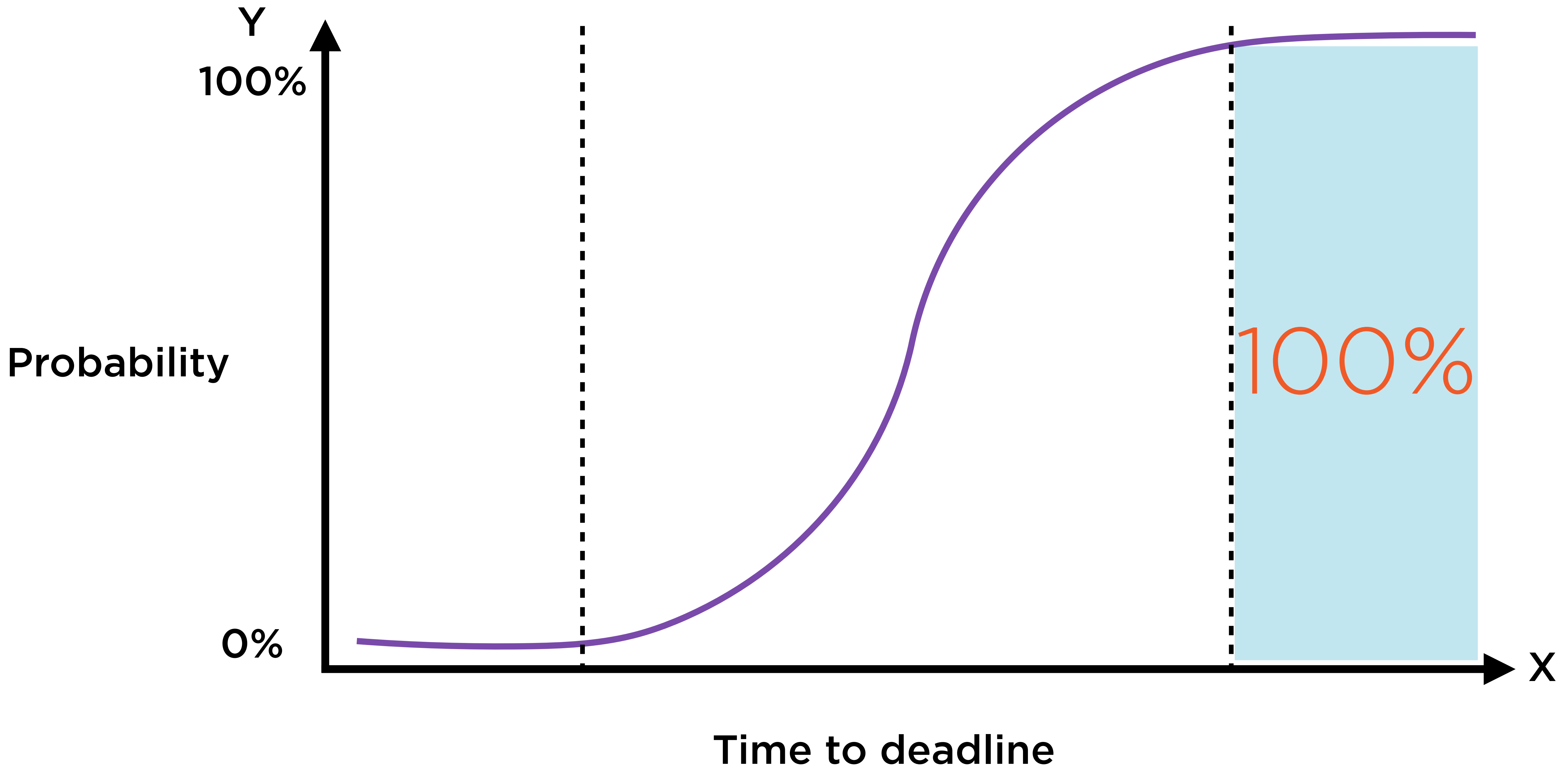


Working Smart with Logistic Regression



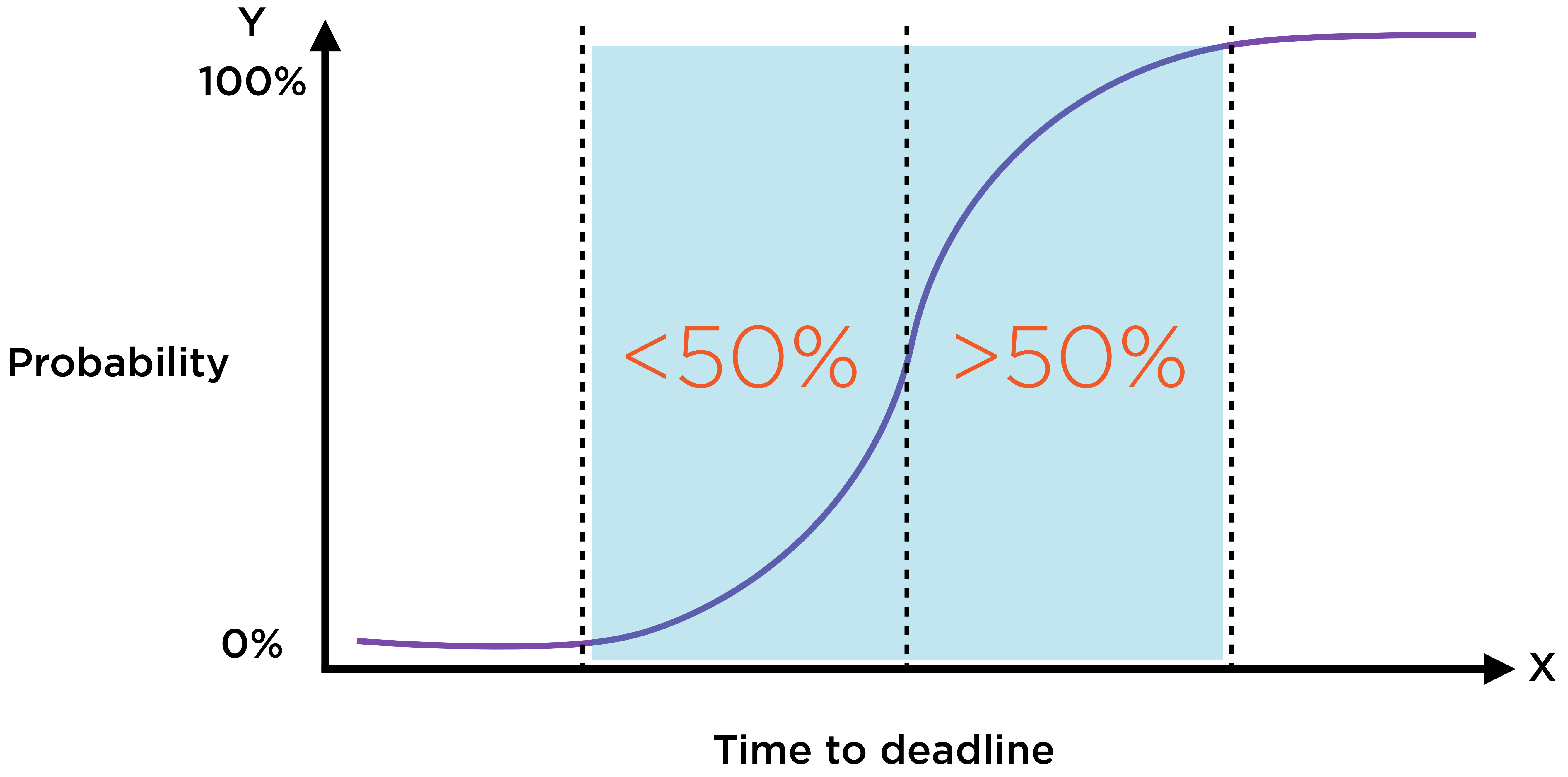
Start too late, and you’ll definitely miss

Working Smart with Logistic Regression



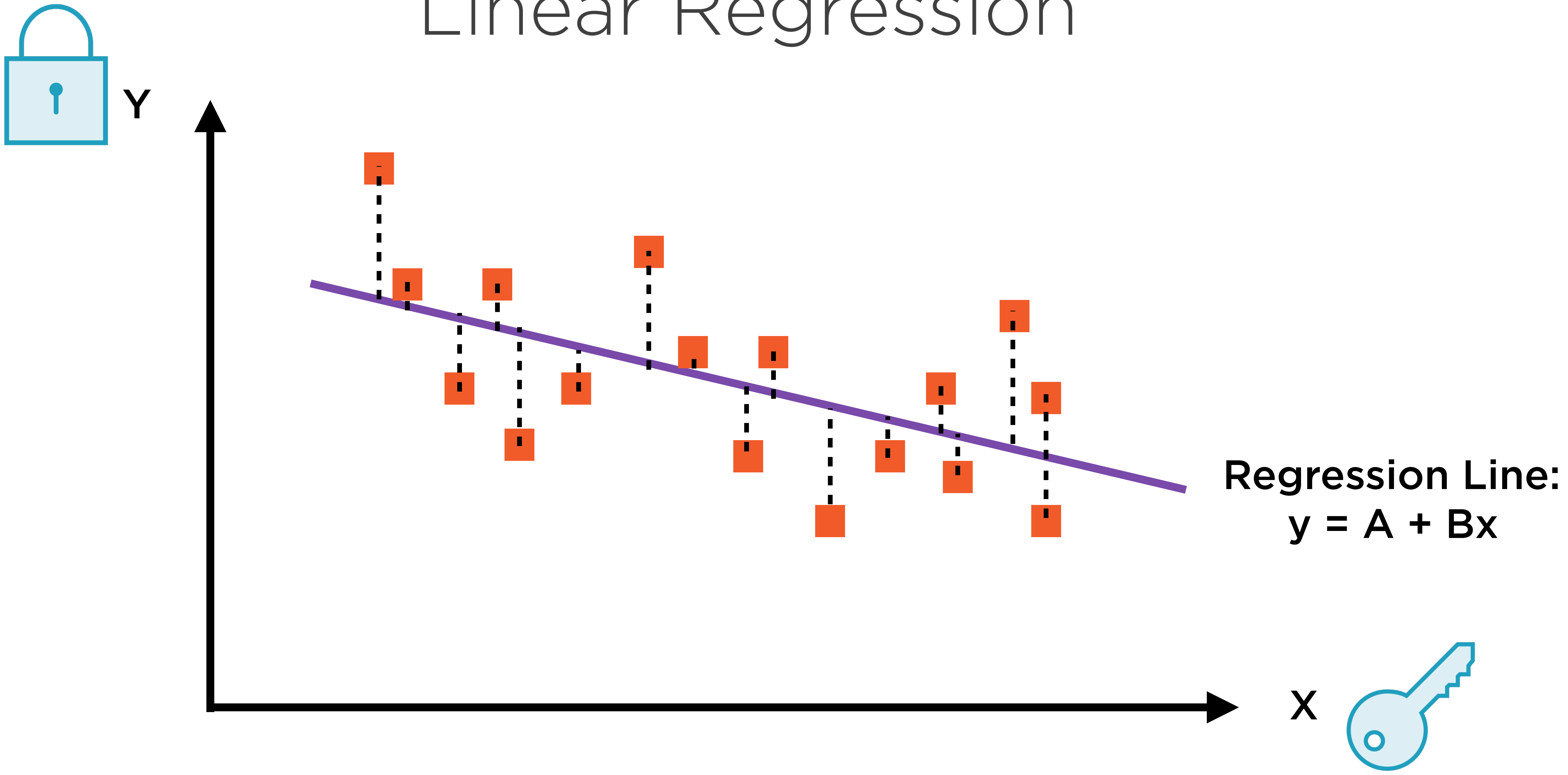
Start too early, and you'll definitely make it

Working Smart with Logistic Regression



Working smart is knowing when to start

Linear Regression

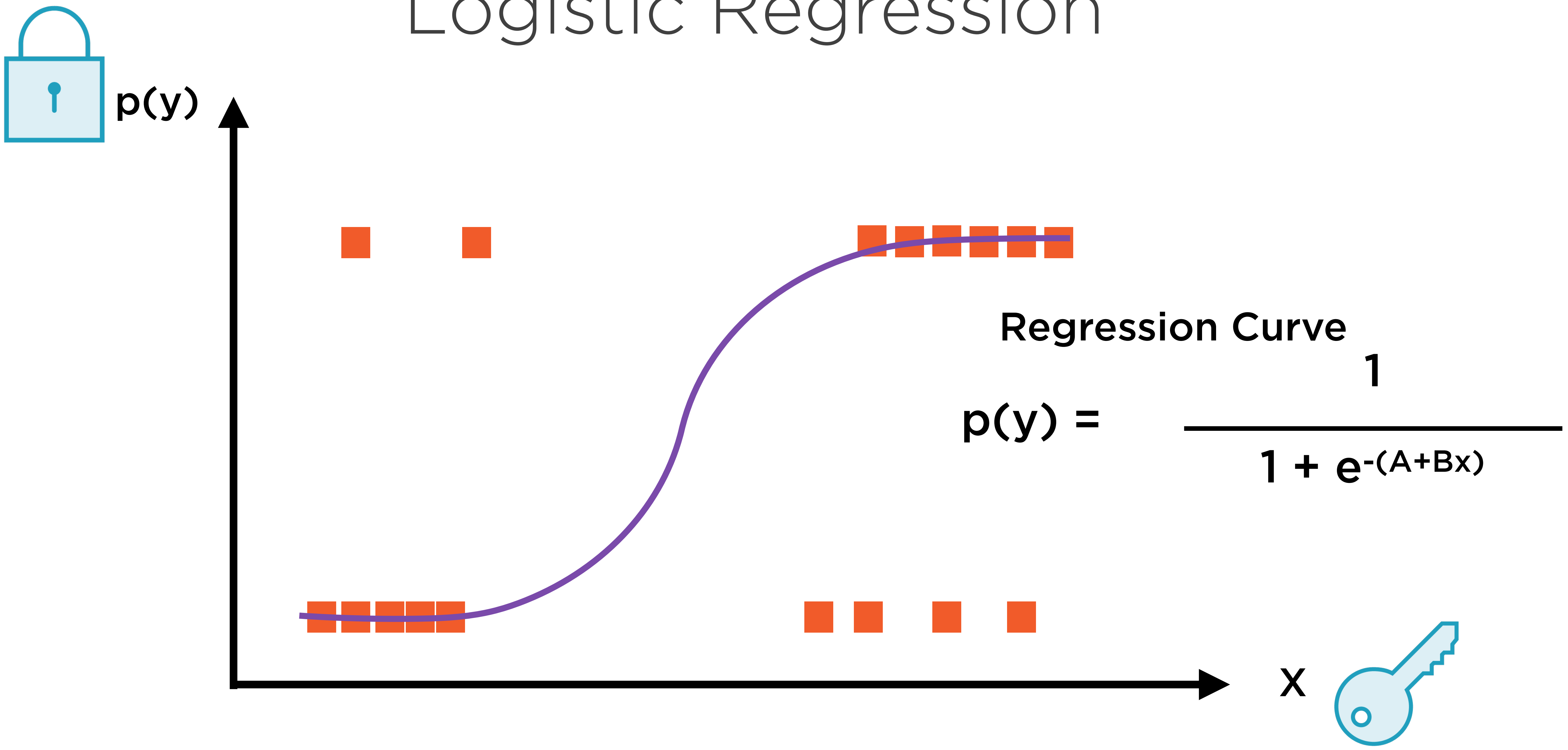


Finding the best fit line through these points

Logistic Regression

Finding the best fit S-curve through these points

Logistic Regression



Finding the best fit S-curve
through these points

Logistic Regression

Regression Equation:

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

Solve for A and B that “best fit” the data

Demo

Building a binary classification model using numeric data

Other Classification Algorithms

Classification Algorithms

Support Vector Machines

Nearest Neighbors

Decision Trees

Naive Bayes

Classification Algorithms

Support Vector Machines

Nearest Neighbors

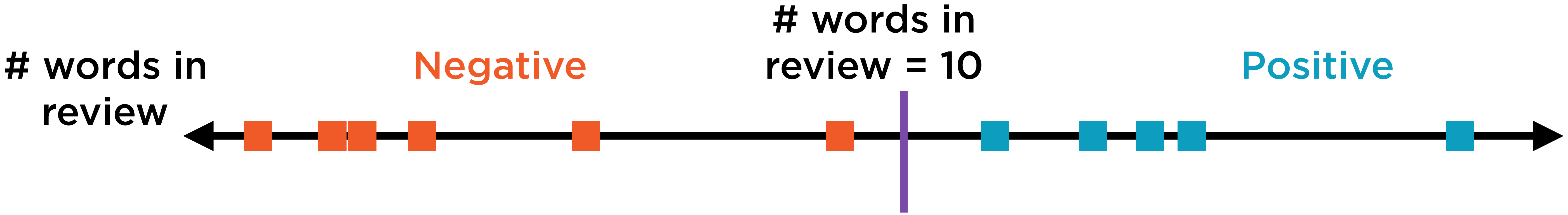
Decision Trees

Naive Bayes

Classify Reviews

Consider data in one dimension

Classify Reviews

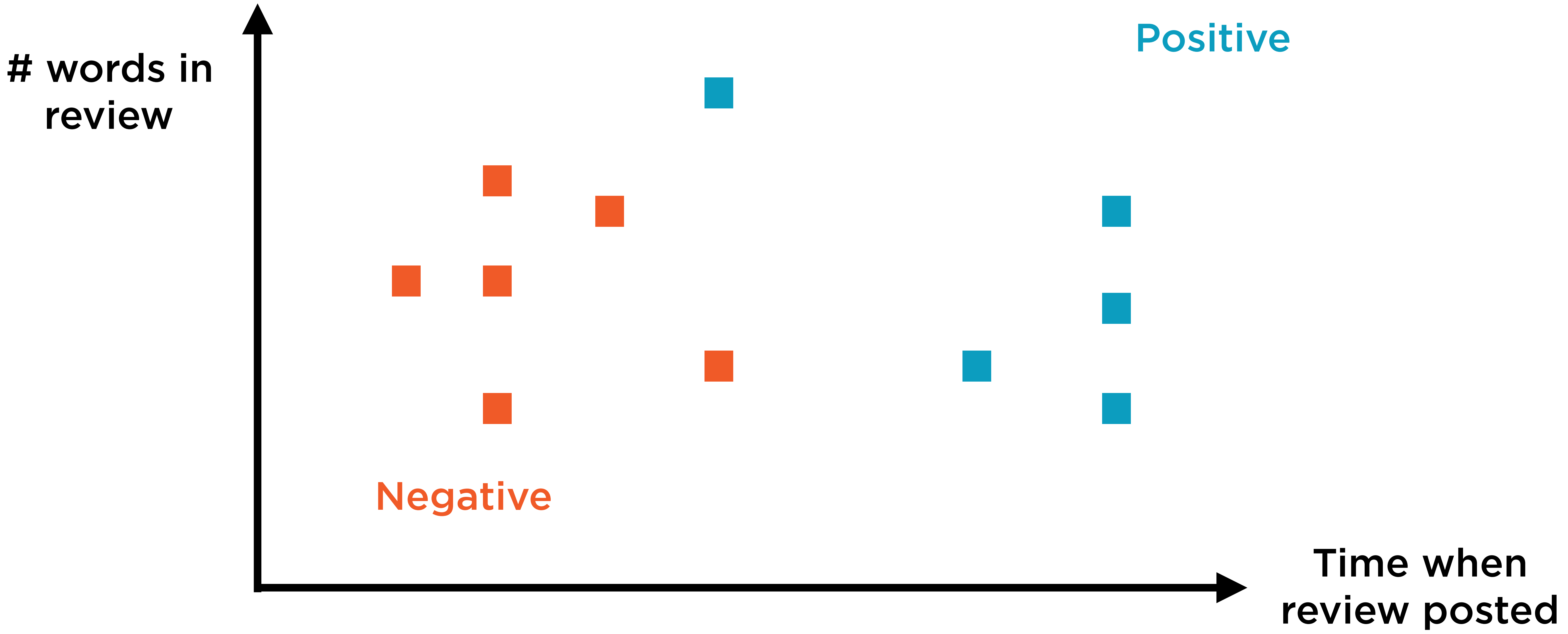


Unidimensional can be separated, or classified, using a point

Classify Reviews

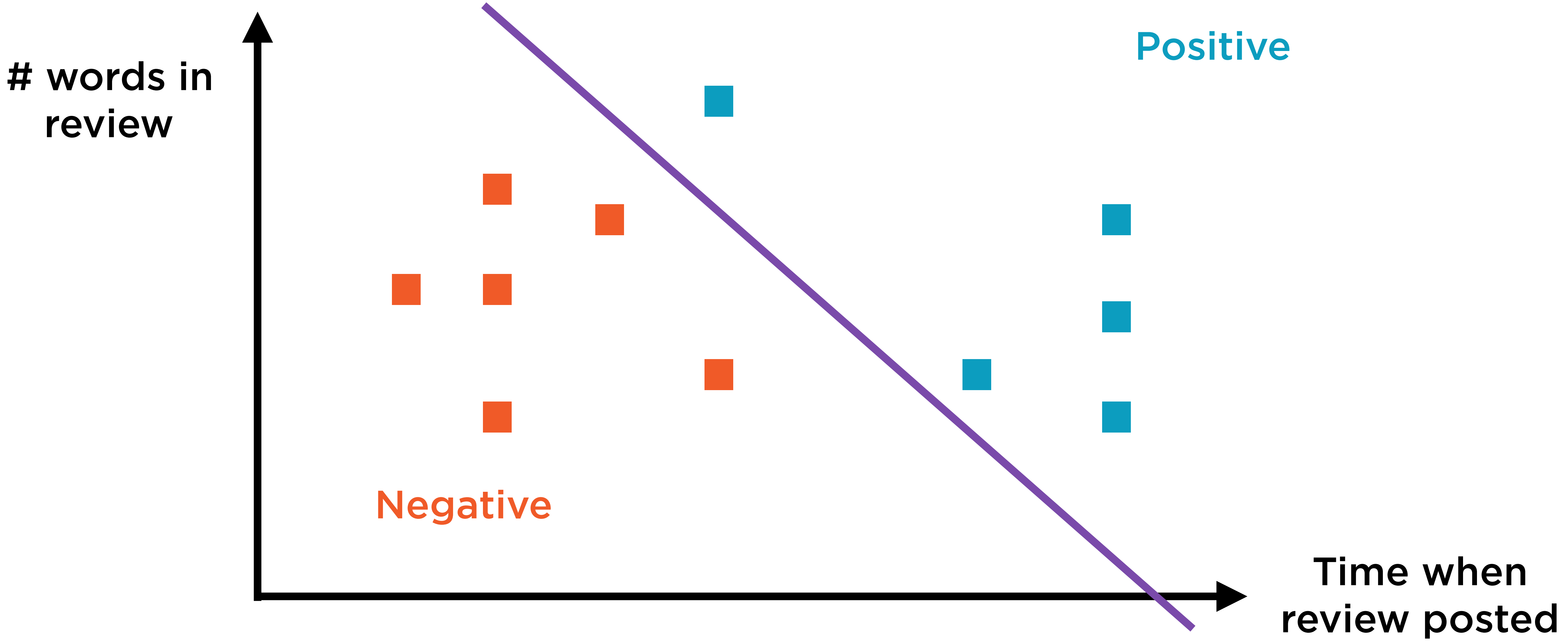
Consider data in two dimensions

Classify Reviews



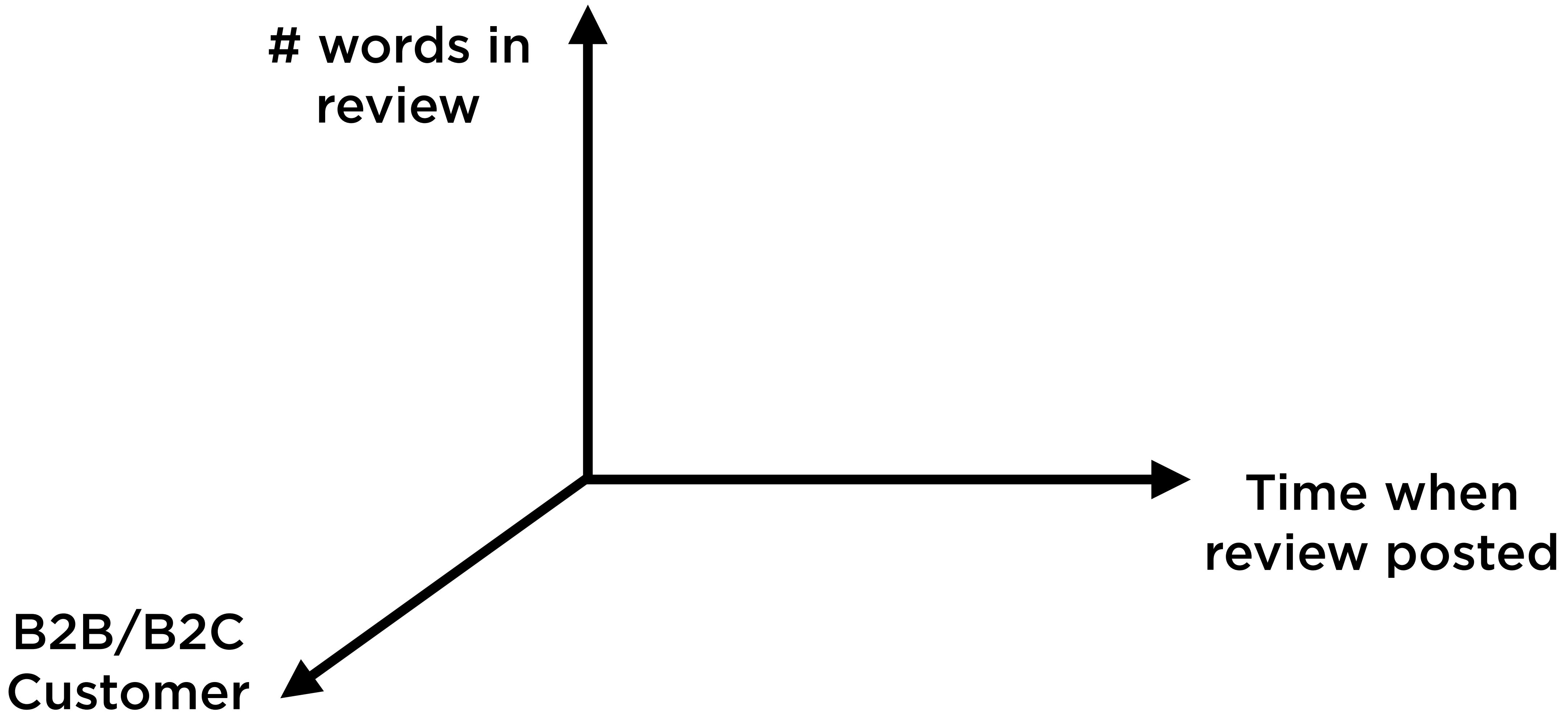
Consider data in two dimensions

Classify Reviews



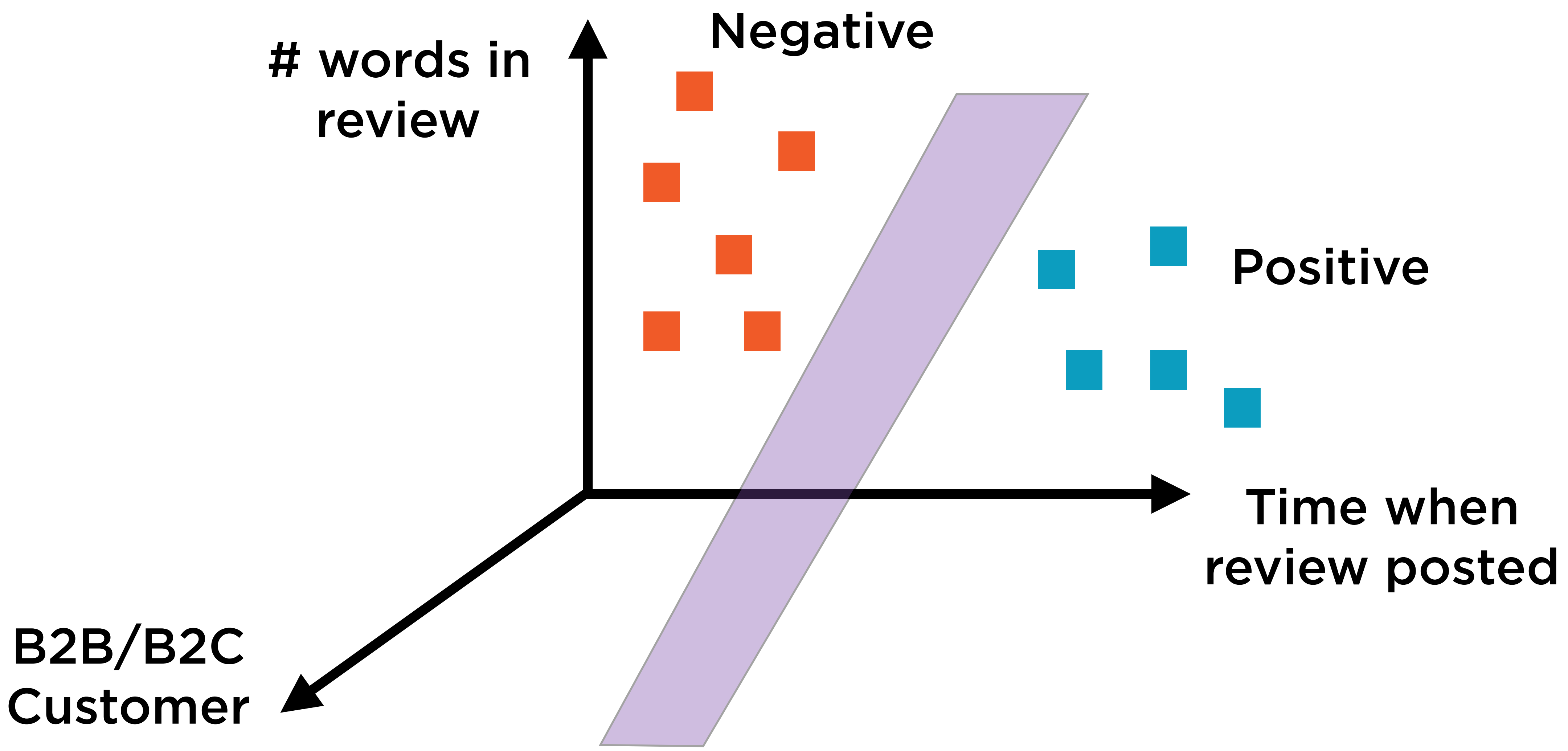
Bidimensional data can be separated, or classified, using a line

Classify Reviews



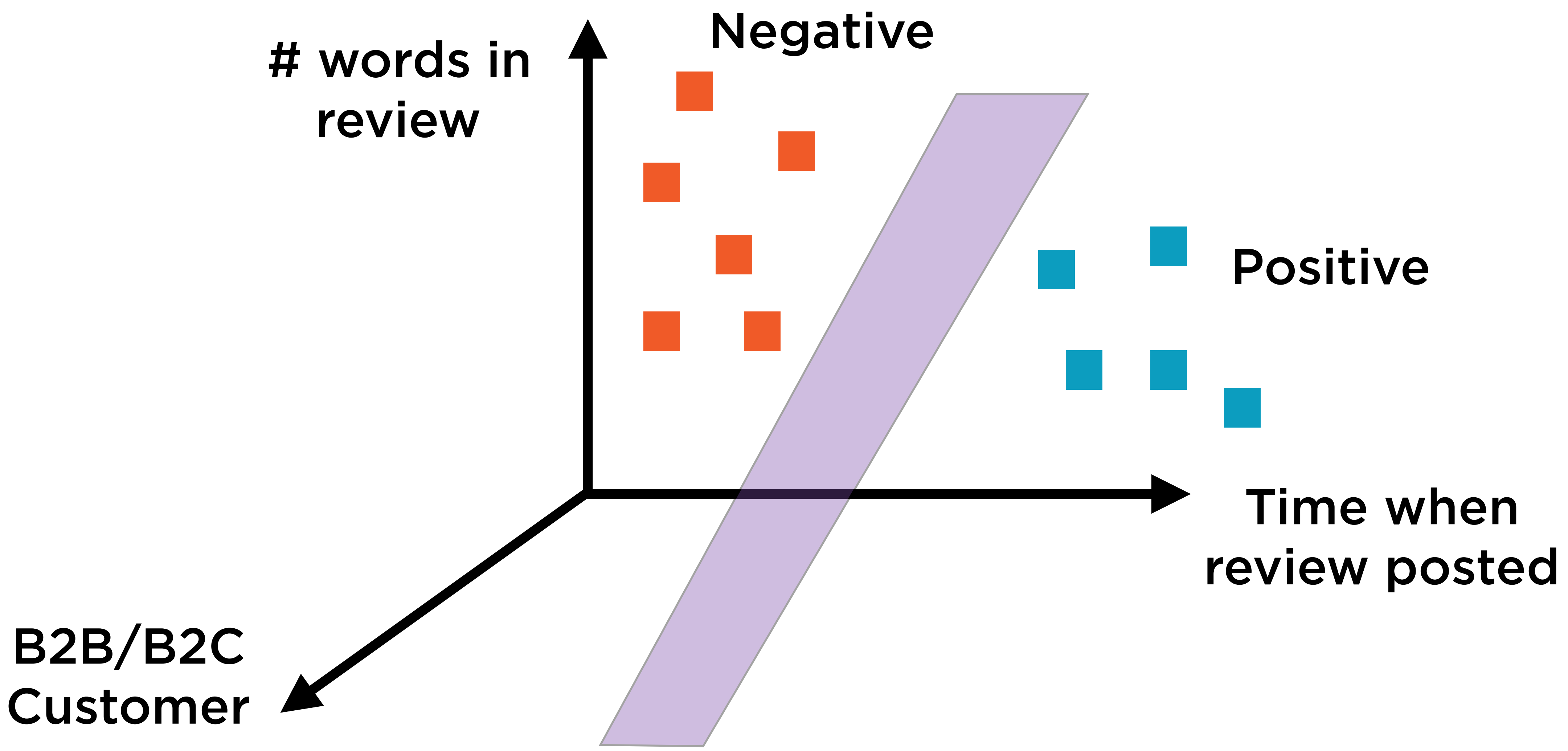
Consider data in 3 dimensions

Classify Reviews



3-dimensional data can be separated, or classified, using a 2-D plane

Classify Reviews

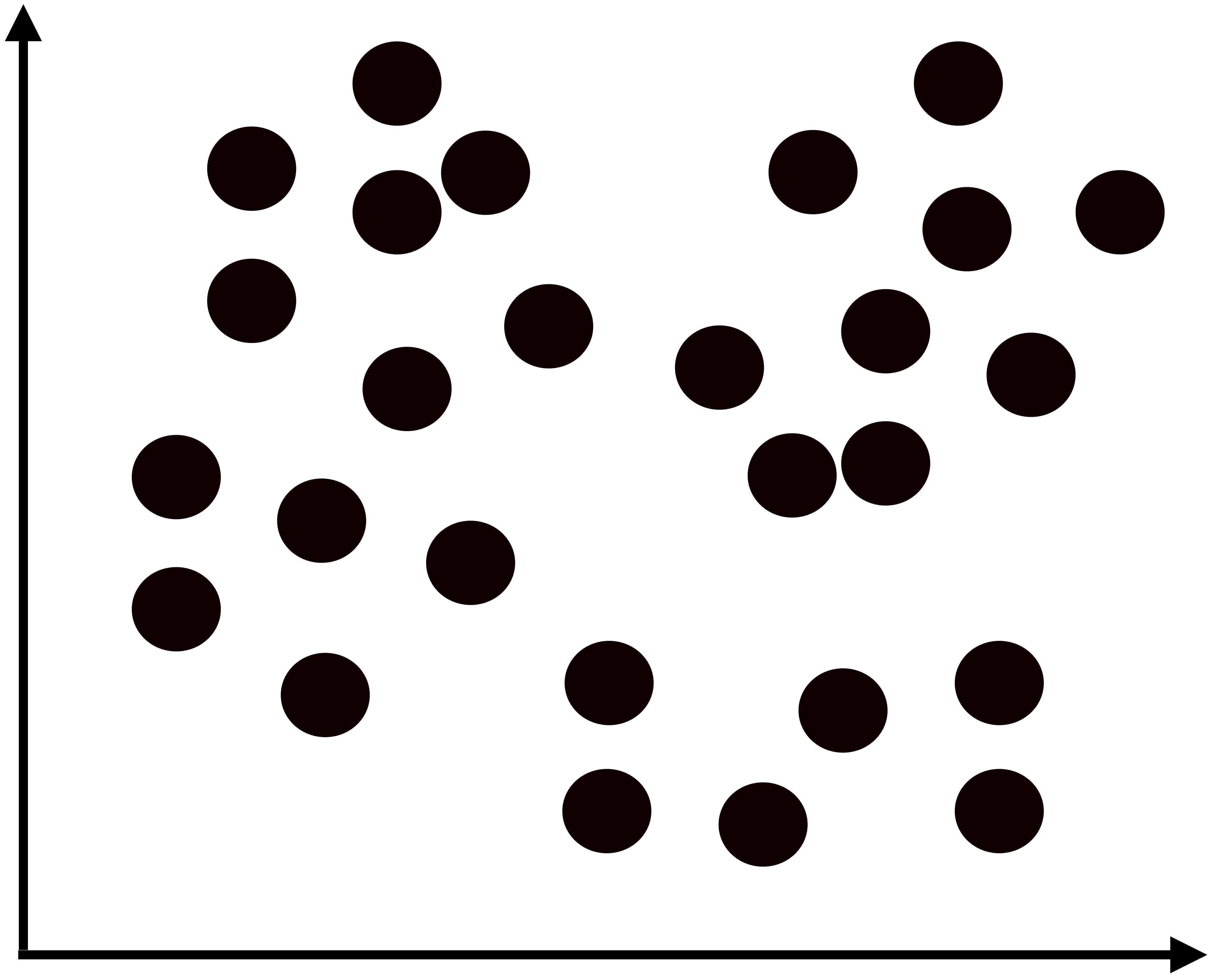


N-dimensional data can be represented in a hypercube, and classified using a hyperplane

Classification Algorithms

Support Vector Machines	**Nearest Neighbors**
Decision Trees	Naive Bayes

Data Points



Nearest Neighbors Classification uses training data to find what is most similar to the current sample

Nearest Neighbors Classification

K-nearest-neighbors Classification

Radius Neighbors Classification

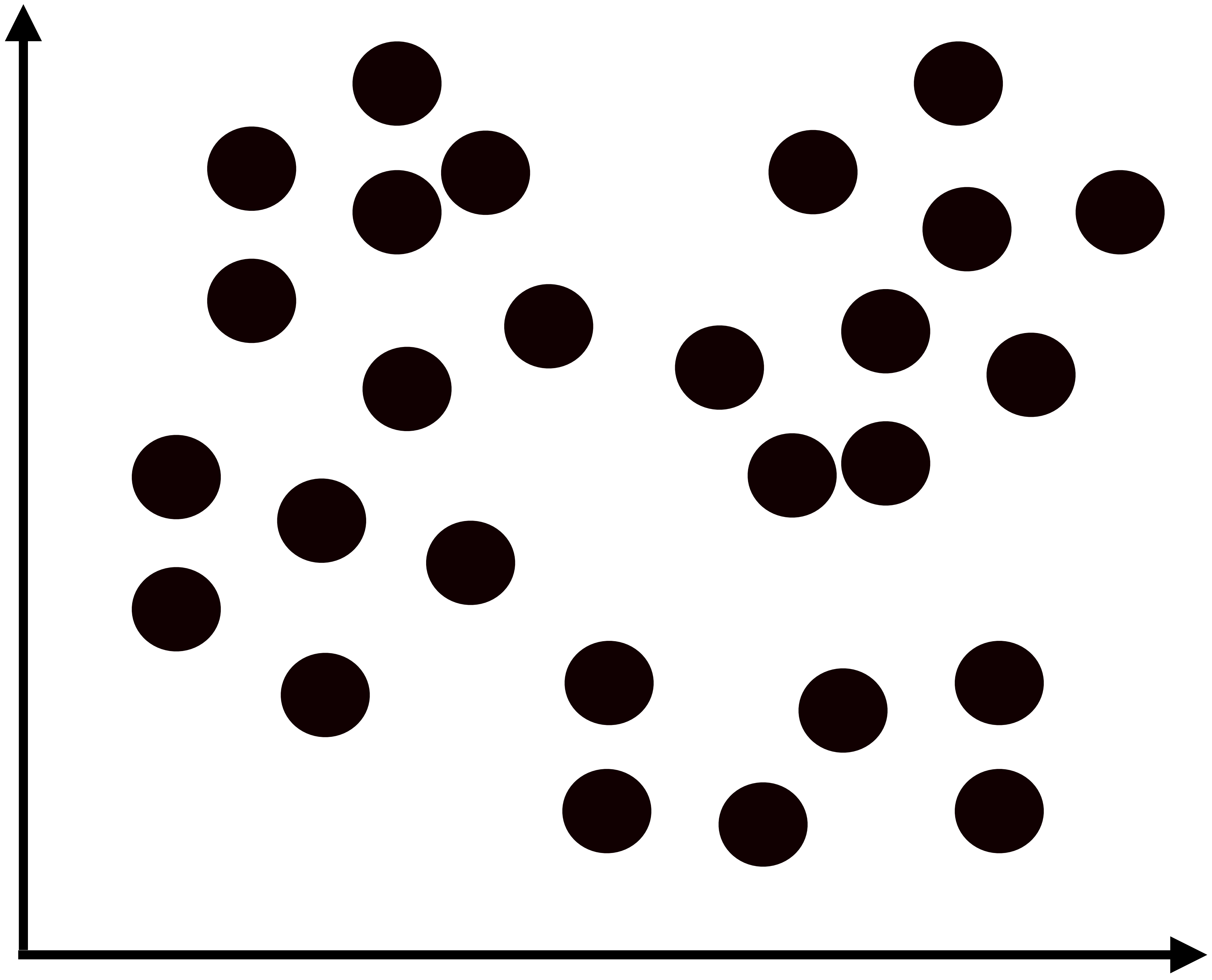
Nearest Neighbors Classification

Voting among K nearest neighbors

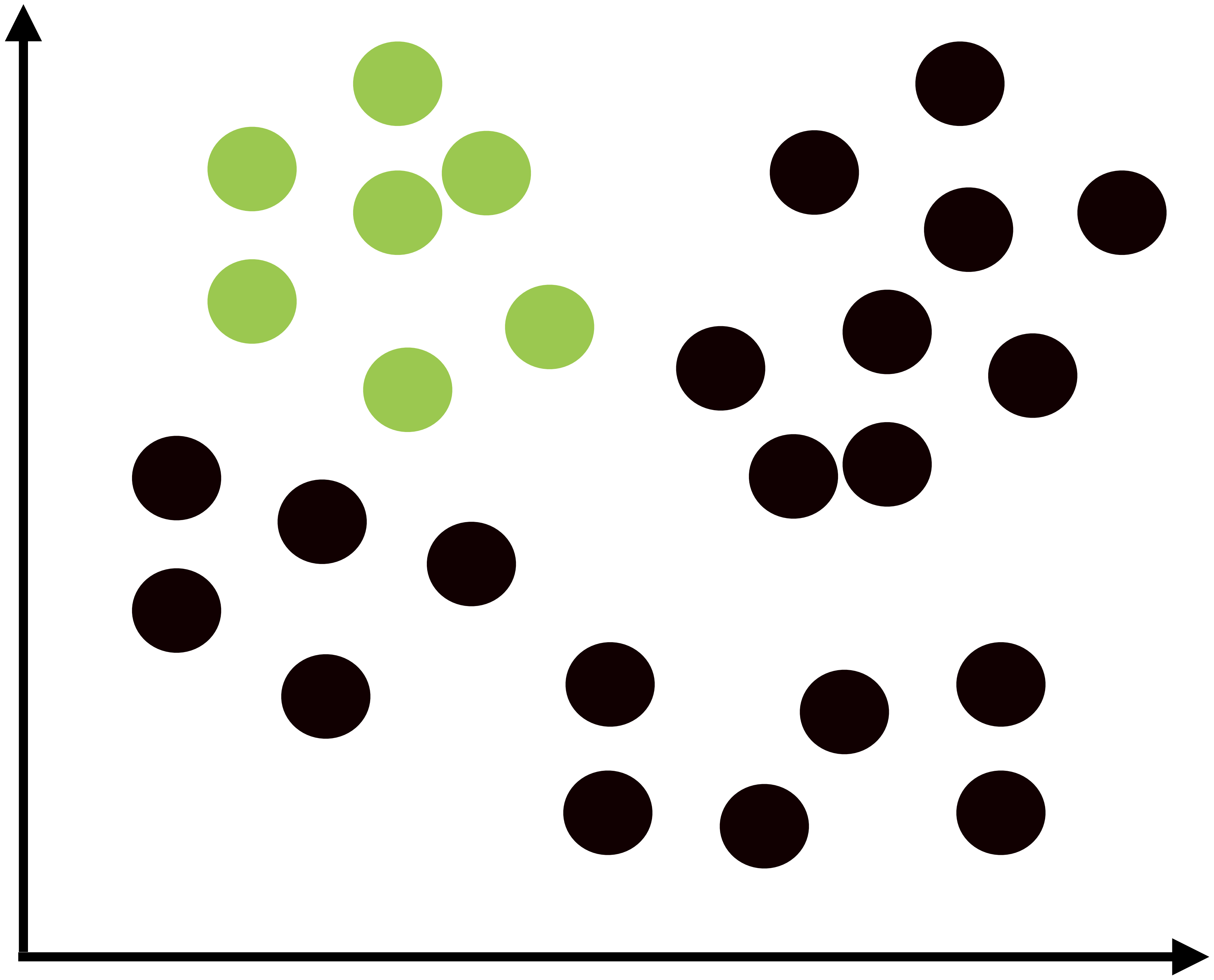
Voting among all neighbors within radius

Classify a data point with the same category as the class to which the neighbors belong

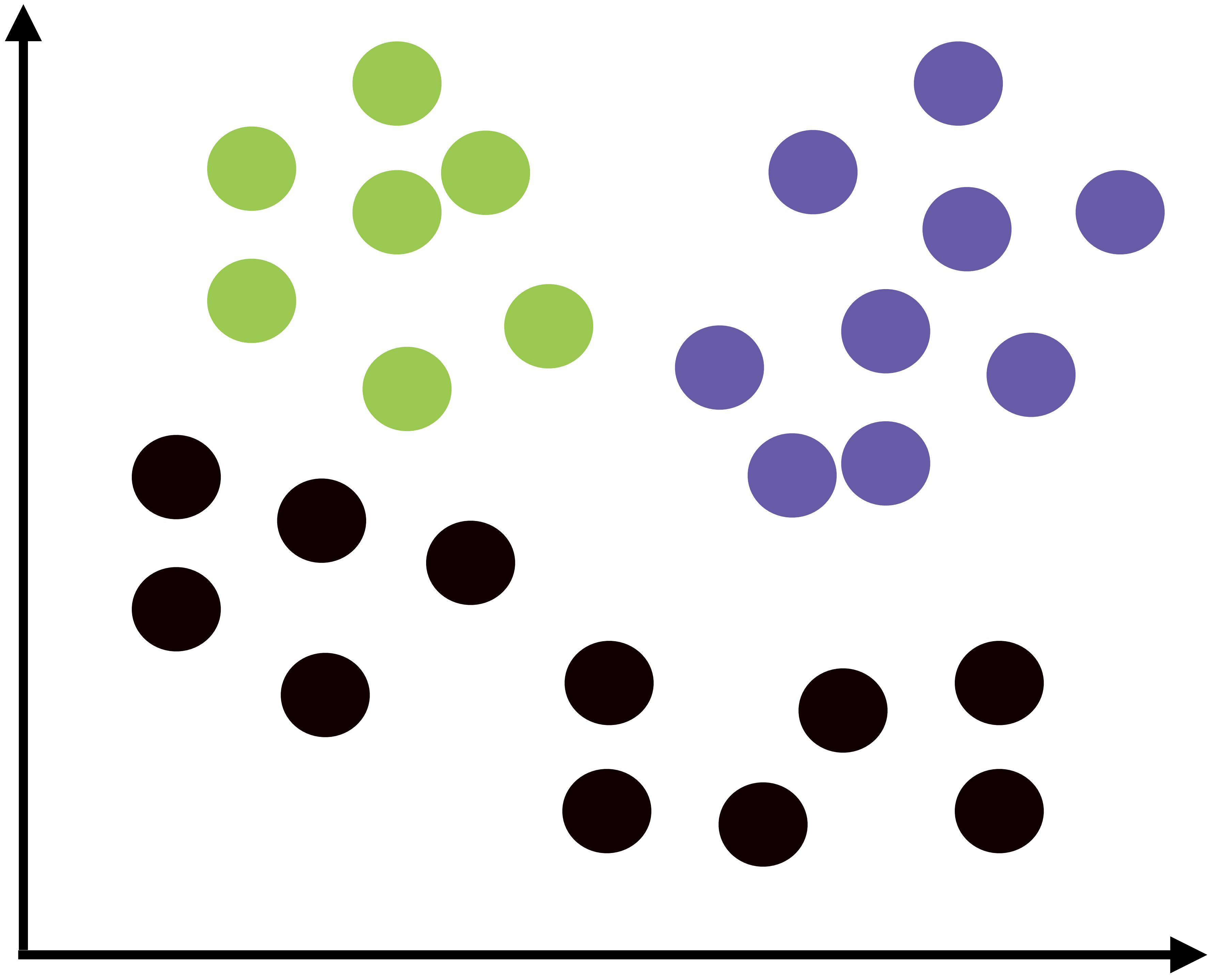
Data Points



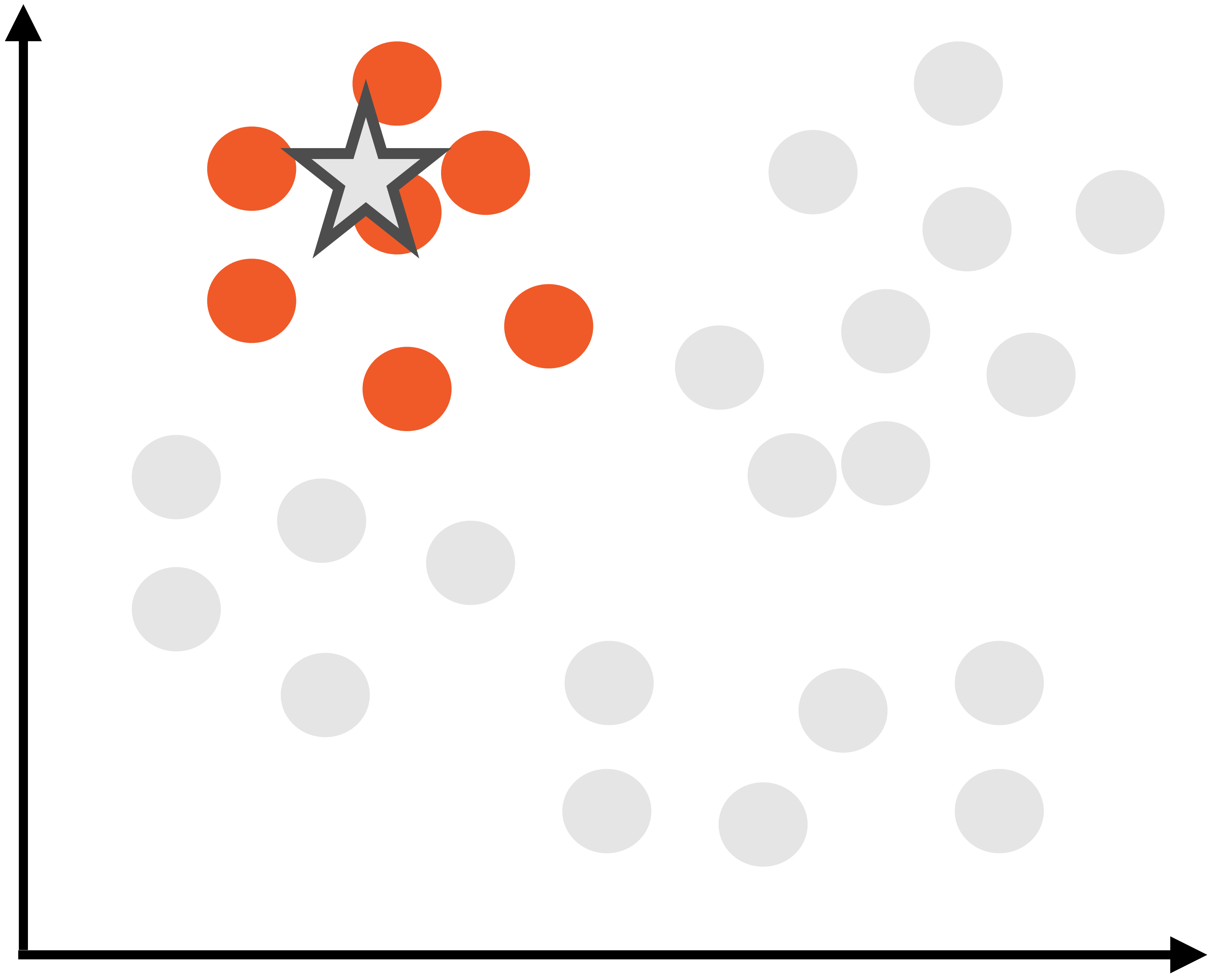
Nearest Neighbors



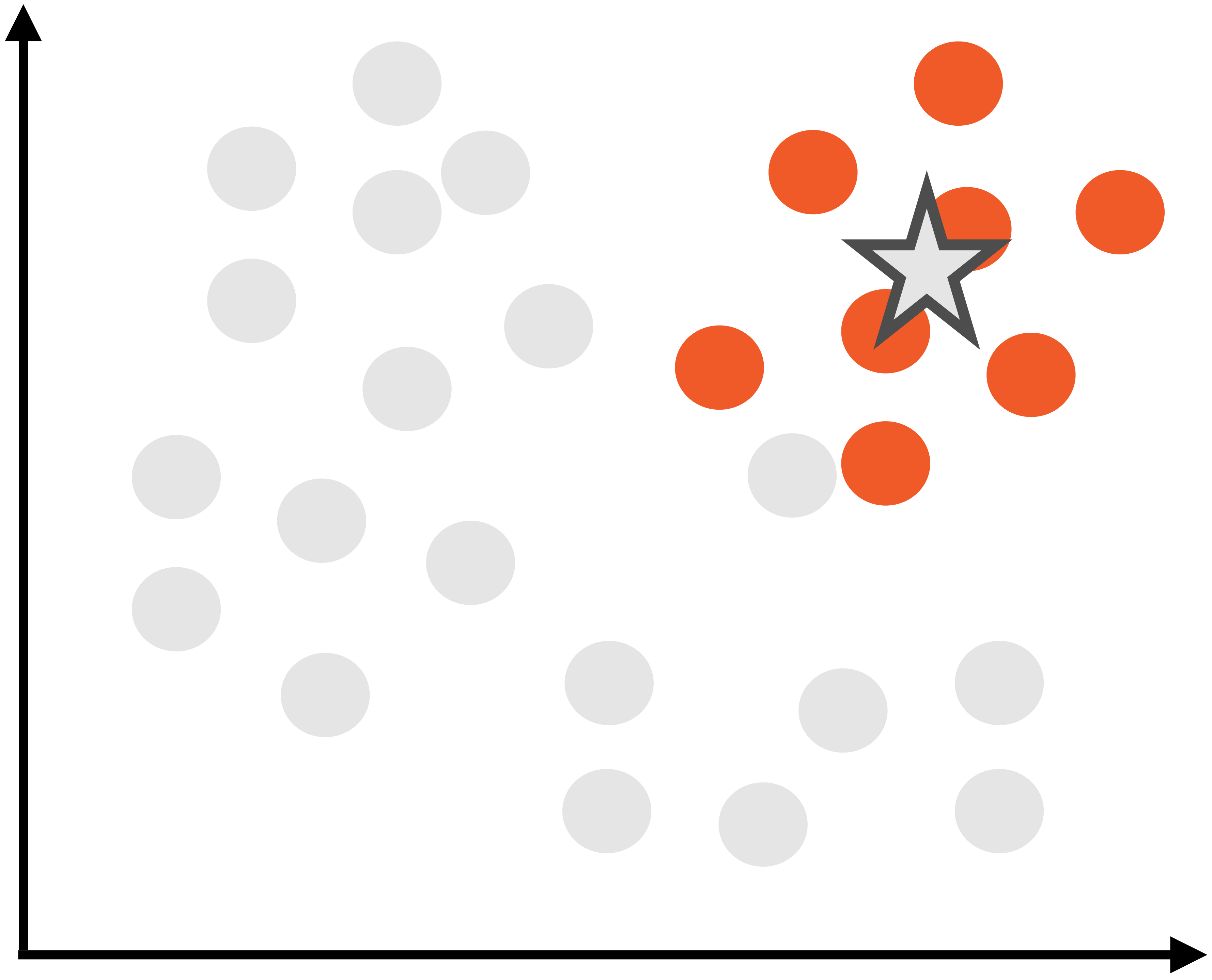
Nearest Neighbors



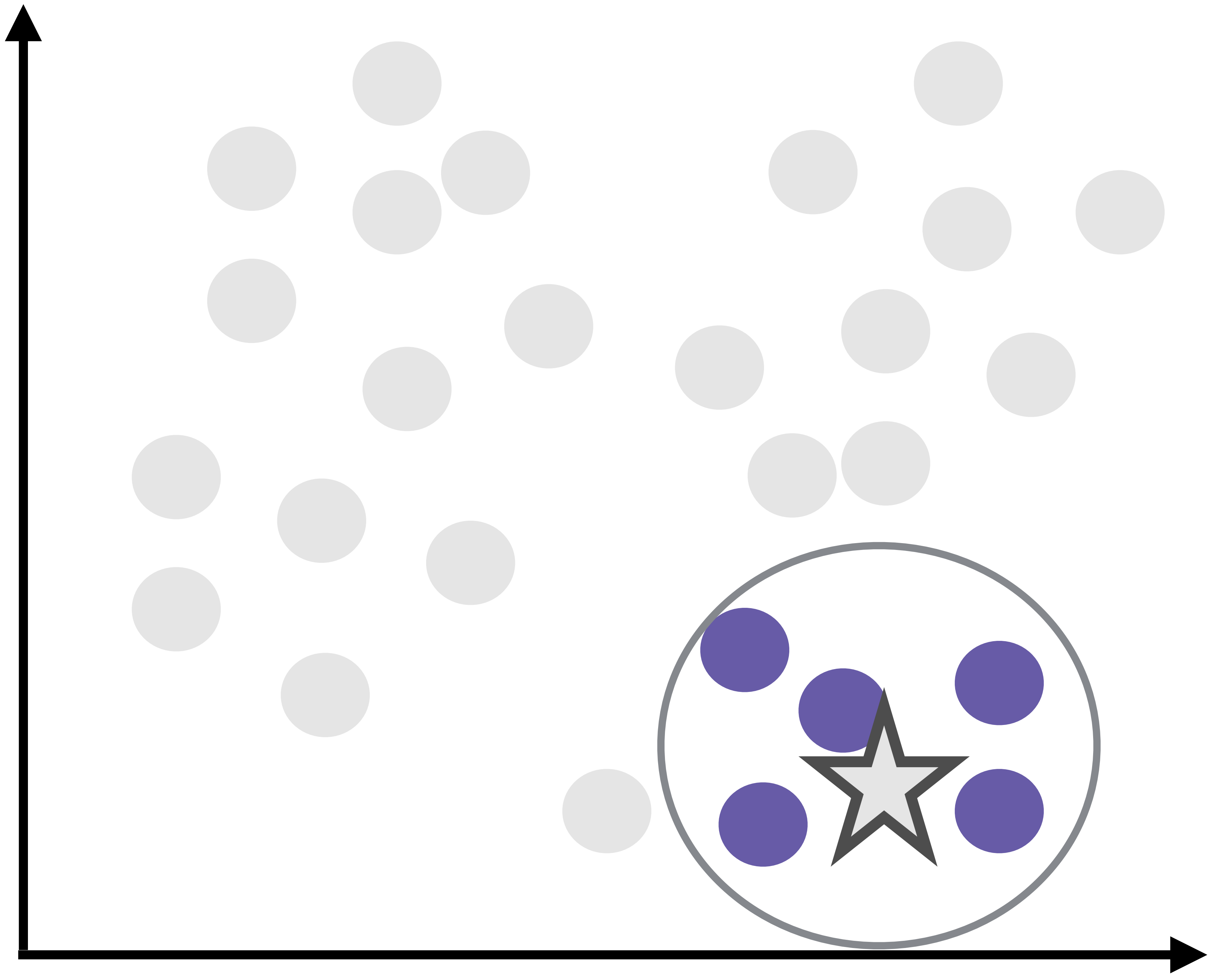
K-nearest-neighbors



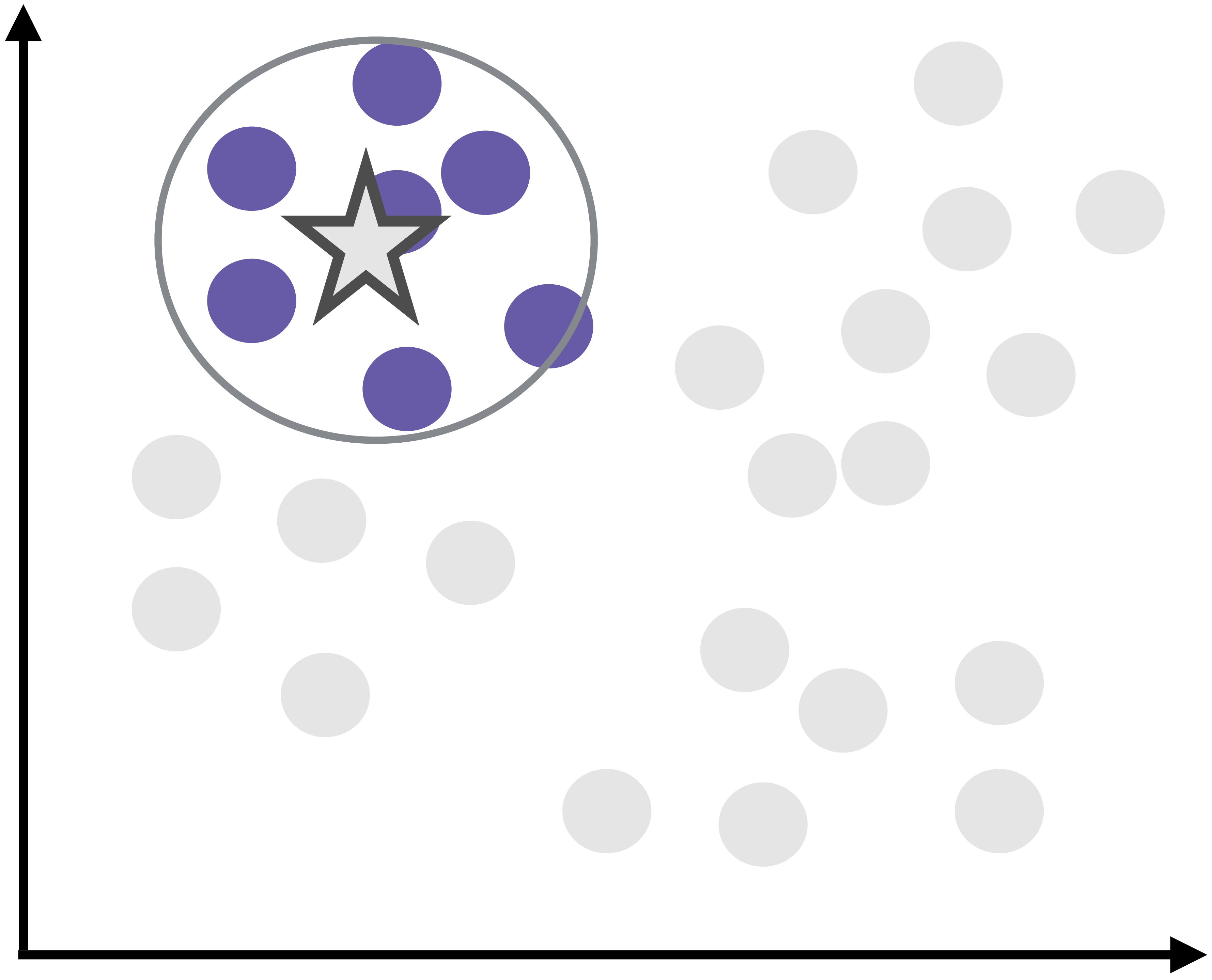
K-nearest-neighbors



Radius Neighbors



Radius Neighbors



Classification Algorithms

Support Vector Machines	Nearest Neighbors
Decision Trees	Naive Bayes

Jockey or Basketball Player?

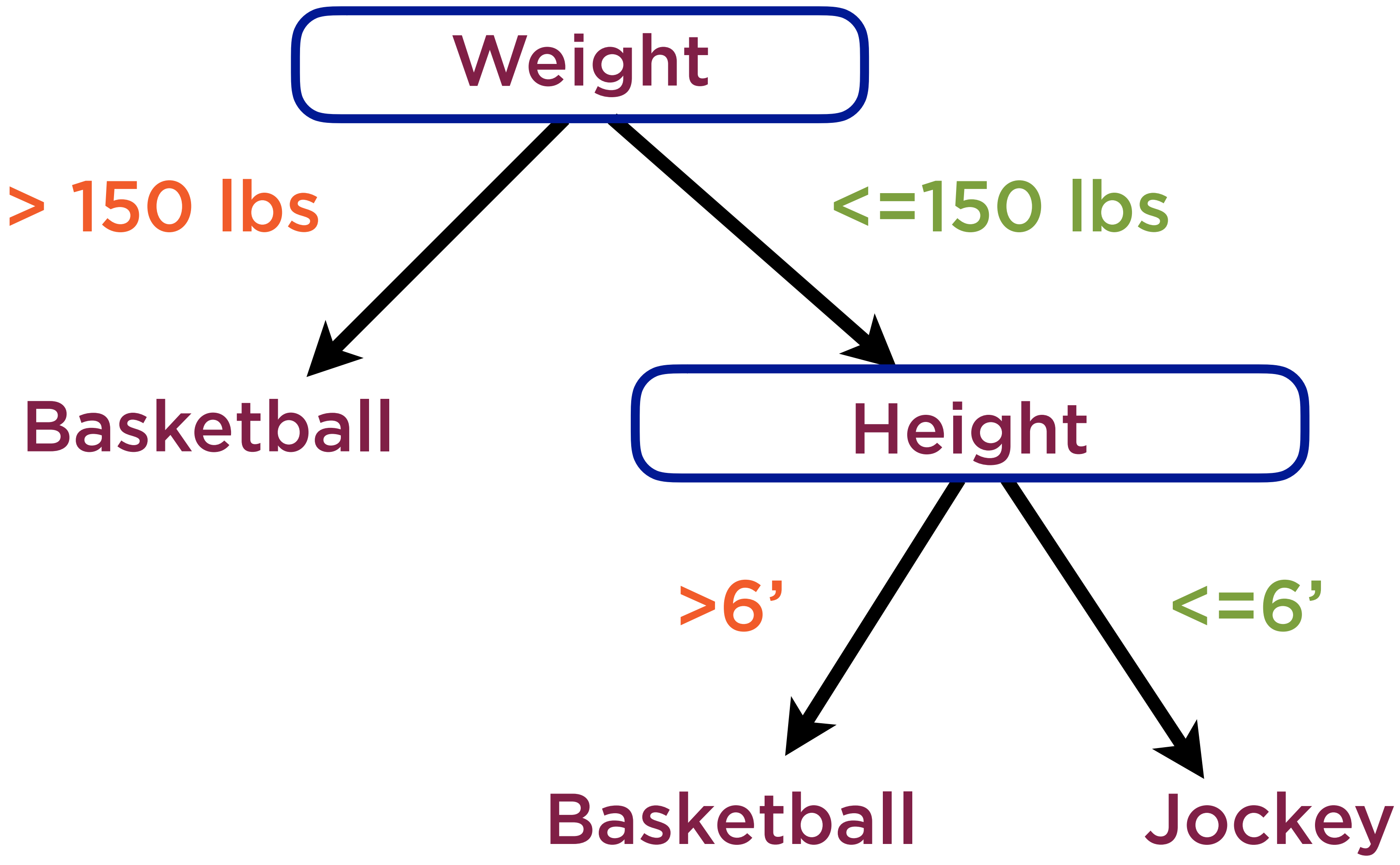
Jockeys

Tend to be light to meet horse carrying limits

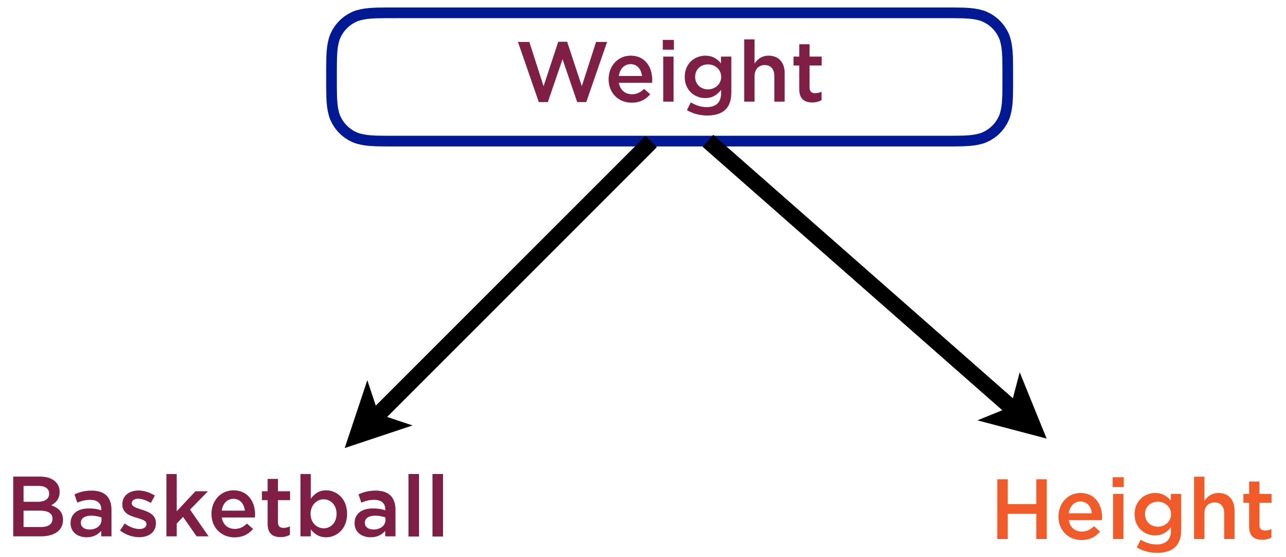
Basketball Players

Tend to be tall, strong and heavy

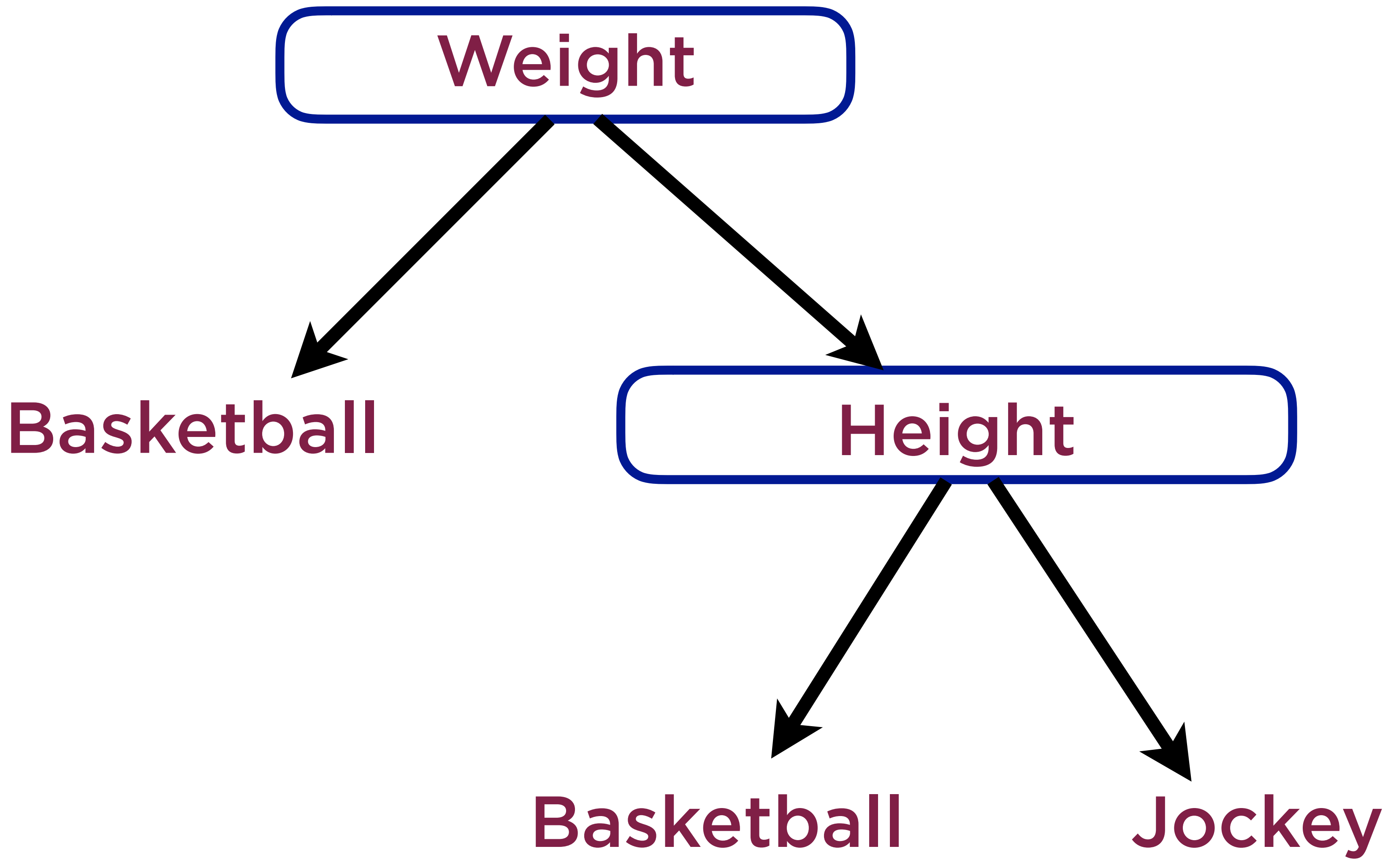
Fit Knowledge into Rules



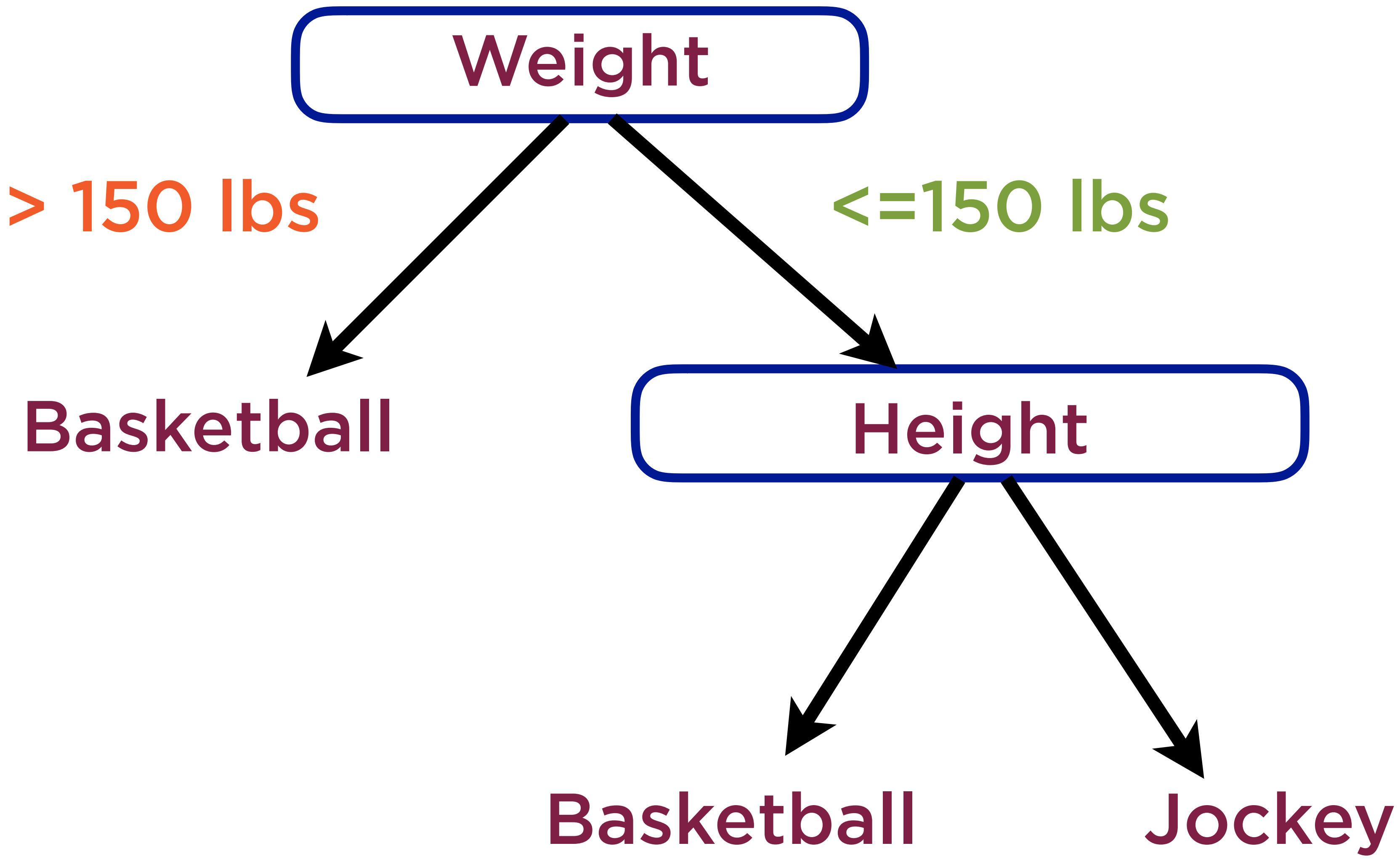
Decision Based on Weight



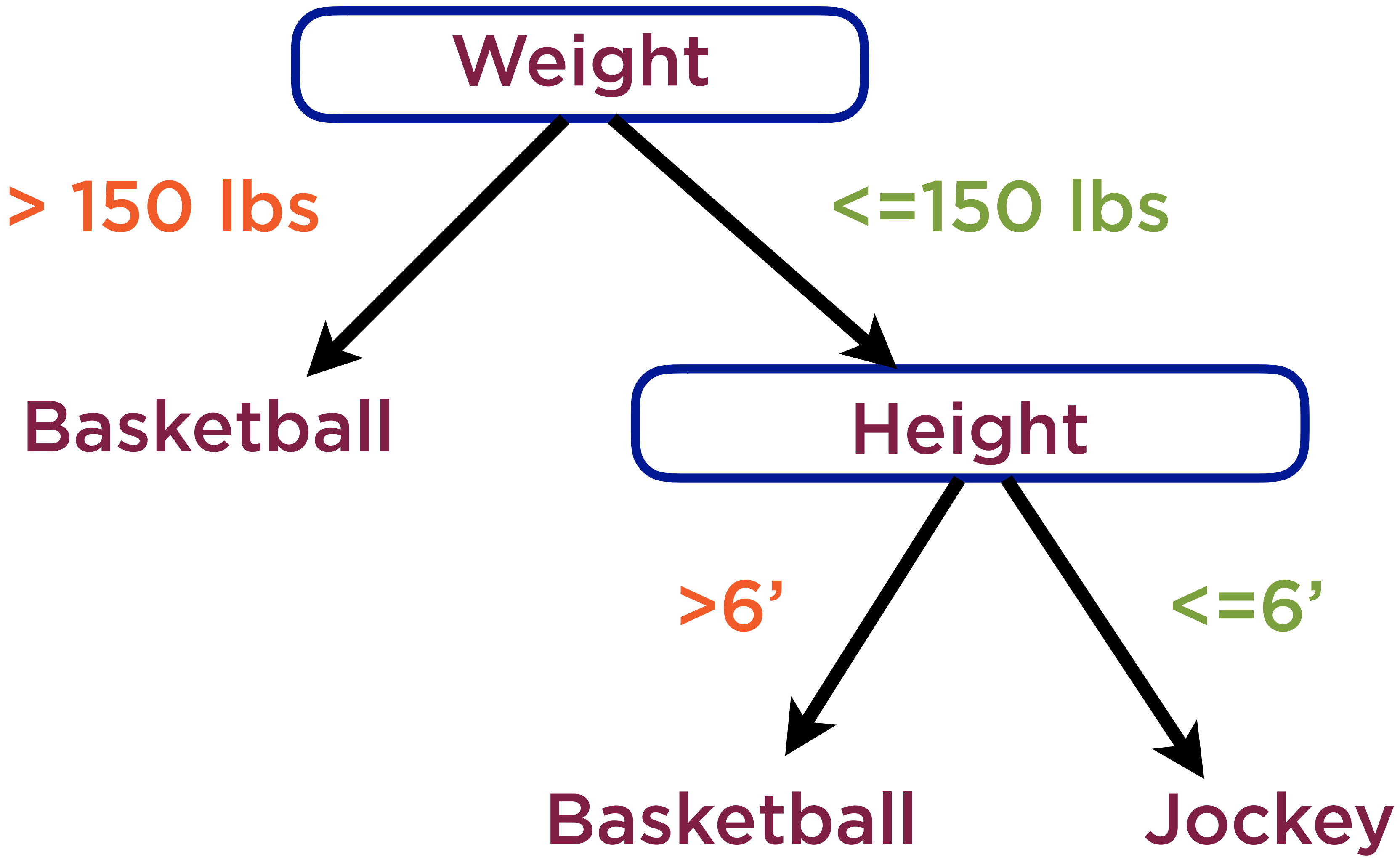
Decision Based on Height



Fit Knowledge into Rules



Fit Knowledge into Rules



Decision trees set up a tree structure on training data which helps make **decisions** based on **rules**

Random Forest

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of training data.

Classification Algorithms

Support Vector Machines

Nearest Neighbors

Decision Trees

Naive Bayes

Binary Classification Problem

Runner

Police Officer

Classify a person who jogs past you on the street

A Priori Probabilities

Items	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

Jogger Is a Police Officer

P(Police Officer/ Handcuffs,Badge) > P(Runner/ Handcuffs,Badge) =

Jogger Is a Marathon Runner

P(Police Officer/
Handcuffs,Badge)

<

P(Runner/
Handcuffs,Badge) =

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

Demo

Performing classification using multiple techniques

Demo

Building an ensemble classifier using warm start

Demo

Performing multiclass classification on text data

Summary

Binary vs. multiclass classification

Logistic regression intuition

Other classification algorithms

Support vector classification

Nearest-neighbors classification

Decision trees for classification

Naive Bayes classification