Machine Learning

Lecture # 1

Introduction to Machine Learning

Machine Learning

A technique that gives **machines** the **ability** to **learn**, without any explicit programming

Machine Learning

In simpler terms, a machine should be able to **see** some **data**, and **learn to make decisions** based on what it has seen

Machine Learning

An example: You are a car dealer, and you have a historical record of **which cars are accident prone**. How can you “teach” a computer to predict which *new* cars will be accident prone?

	Maximum Speed	Acceleration	Color	Car age	Accident Prone?
Car 1	240 km/h	Fast	Red	2 yrs	Yes
Car 2	100 km/h	Fast	Yellow	2 yrs	No
Car 3	240 km/h	Fast	Blue	1 yr	No
Car 4	200 km/h	Slow	Blue	5 yrs	Yes
Car 5	100 km/h	Fast	Yellow	5 yrs	Yes
Car 6	100 km/h	Slow	Black	6 yrs	No
Car 7	150 km/h	Fast	Red	2 yrs	?

Machine Learning

How can you make your decision?

Search for closest vehicle in the past?

Come up with a set of rules?

How can you decide what knowledge is important?

Machine Learning

Historically, rule based systems were common:

```
if (car.acceleration = fast and car.age > 1 and . . . )
    print (“accident prone”)
else if (car.acceleration = slow and car.maxspeed > 150 and . . . )
    print (“accident prone”)
else if (car.acceleration = slow and car.maxspeed < 50)
    print (“not accident prone”)
else if
 . . .
 . . .
```

Domain Specific

Cumbersome

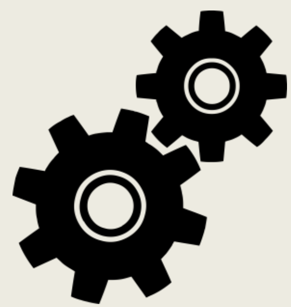
Not easy to learn from new data

Machine Learning

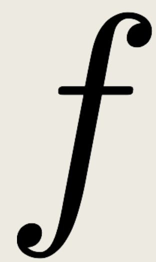
Then, machine learning techniques came about...

	...	Acciden t Prone?
Car 1	...	Yes
Car 2	...	No
Car 3	...	No
Car 4	...	Yes
Car 5	...	Yes
Car 6	...	No

Training data



Machine Learning Algorithms



“Model”

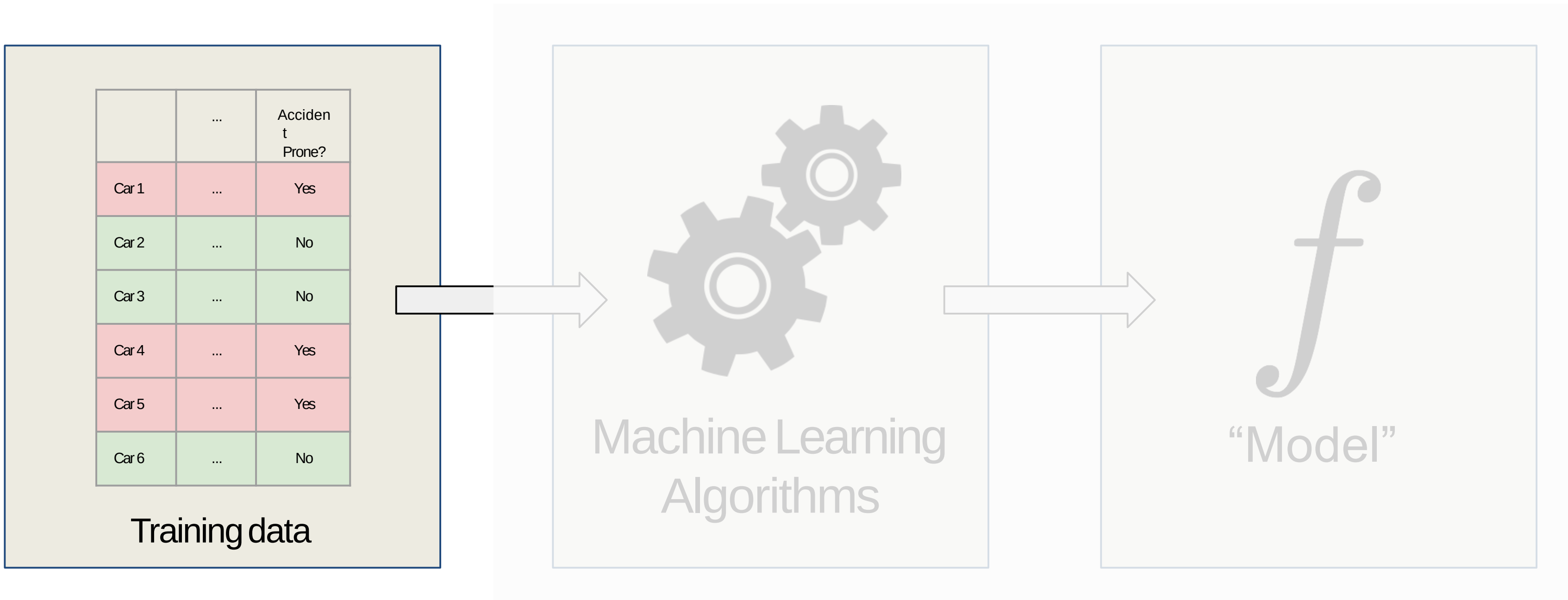
Domain Agnostic

Robust

Easy to learn from new data

Machine Learning

Let’s talk about training data



Domain Agnostic

Robust

Easy to learn from new data

Training Data

	Maximum Speed	Acceleration	Color	Car age	Accident Prone?
Car 1	240 km/h	Fast	Red	2 yrs	Yes
Car 2	100 km/h	Fast	Yellow	2 yrs	No
Car 3	240 km/h	Fast	Blue	1 yr	No
Car 4	200 km/h	Slow	Red	5 yrs	Yes
Car 5	100 km/h	Fast	Yellow	5 yrs	Yes
Car 6	100 km/h	Slow	Black	6 yrs	No

Input Features Labels

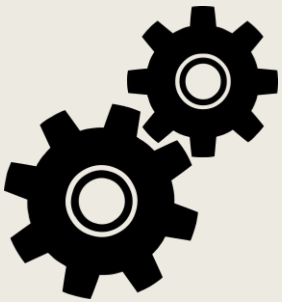
We use training examples with labels to train a model

Machine Learning

Algorithms use this training data

	...	Accident Prone?
Car 1	...	Yes
Car 2	...	No
Car 3	...	No
Car 4	...	Yes
Car 5	...	Yes
Car 6	...	No

Training data



Machine Learning Algorithms

f

"Model"

Domain Agnostic

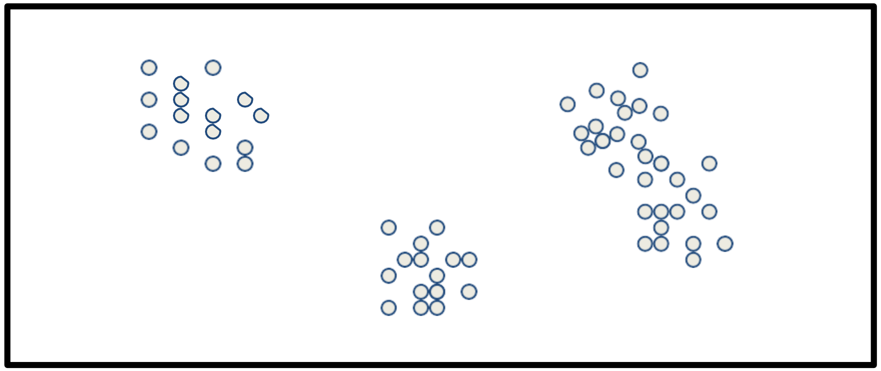
Robust

Easy to learn from new data

Algorithms

In this case, we have labels for each car. This class of problems is handled by *supervised learning algorithms*.

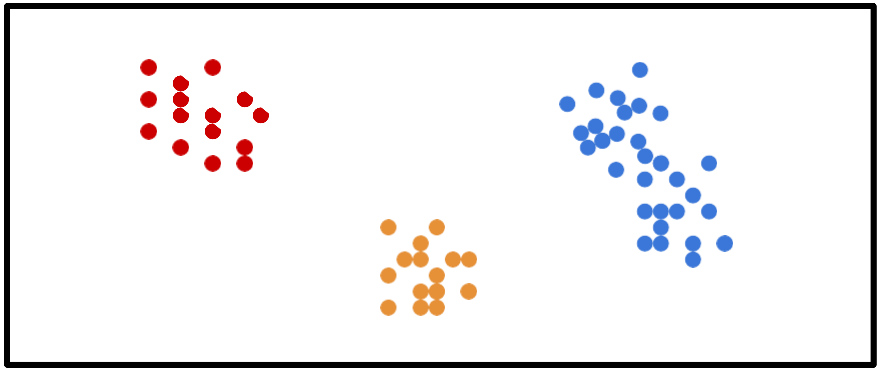
Unsupervised learning algorithms work on unlabelled data



Algorithms

In this case, we have labels for each car. This class of problems is handled by *supervised learning algorithms*.

Unsupervised learning algorithms work on unlabelled data



Algorithms

In this case, we have labels for each car. This class of problems is handled by *supervised learning algorithms*.

Unsupervised learning algorithms work on unlabelled data

Algorithms

Many techniques exist to build models:

- “Finding similar cars” type methods:
  - K-means clustering
  - Hierarchical clustering
- “Create set of rules” type methods:
  - Support vector machines
  - Logistic Regression
  - Neural Networks

Algorithms

Many techniques exist to build models:

- "Finding similar cars" type methods:
  - K-means clustering
  - Hierarchical clustering

Unsupervised

- "Create set of rules" type methods:
  - Support vector machines
  - Logistic Regression
  - Neural Networks

Supervised

Machine Learning

Algorithms use this training data to produce a model

	...	Accident Prone?
Car 1	...	Yes
Car 2	...	No
Car 3	...	No
Car 4	...	Yes
Car 5	...	Yes
Car 6	...	No

Training data

Machine Learning Algorithms

f

"Model"

Domain Agnostic

Robust

Easy to learn from new data

Machine Learning Model

Example of a model: *A function that takes information about a car, and predicts whether it's accident prone or not*

f() = No

f () = Yes

Machine Learning Model

Example of a model: *A function that takes a word, and predicts it's part of speech tag*

f("car") = Noun

f("beautiful") = Adjective

f("she") = Pronoun

Machine Learning Model

At a high level, the basic idea is to figure out which features are important, and how important are they for prediction

```
  0.3  x  car.maxspeed
+ 0.2  x  car.acceleration
+ 0.0  x  car.color
+ 0.5  x  car.age
```

Machine Learning Model

At a high level, the basic idea is to figure out which features are important, and how important are they for prediction

```
  0.3 x car.maxspeed
+ 0.2 x car.acceleration
+ 0.0 x car.color
+ 0.5 x car.age
```

An older car is more likely to be accident prone

Machine Learning Model

At a high level, the basic idea is to figure out which features are important, and how important are they for prediction

```
   0.3 x car.maxspeed
 + 0.2 x car.acceleration
 + 0.0 x car.color
 + 0.5 x car.age
```

The color of a car has no impact on accidents

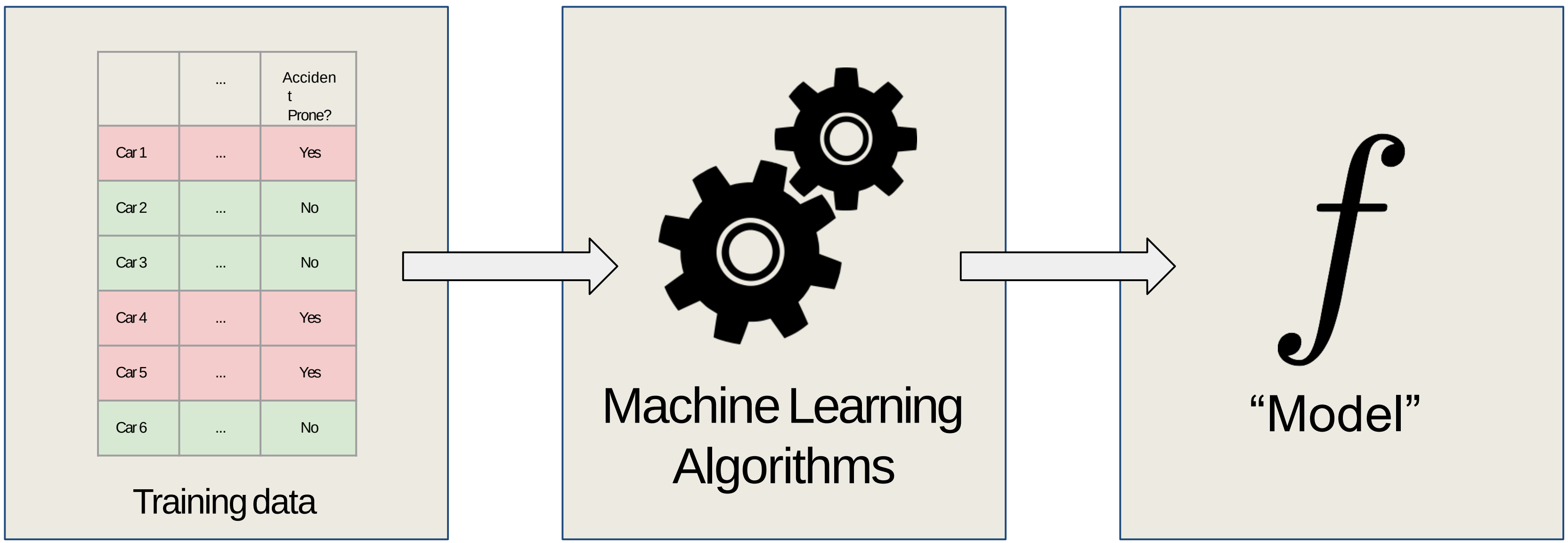
Machine Learning Model

At a high level, the basic idea is to figure out which features are important, and how important are they for prediction

$$
\begin{aligned}
& \boxed{0.3} \times \text{car.maxspeed} \\
+ & \boxed{0.2} \times \text{car.acceleration} \\
+ & \boxed{0.0} \times \text{car.color} \\
+ & \boxed{0.5} \times \text{car.age}
\end{aligned}
$$

Feature Weights

Machine Learning



Supervised Learning: **Classification**

Process of assigning objects to categories

For example, Car 1 belongs to category *“Accident Prone”*

	Maximum Speed	Acceleration	Color	Car age	Accident Prone?
Car 1	240 km/h	Fast	Red	2 yrs	Yes
Car 2	100 km/h	Fast	Yellow	2 yrs	No
Car 3	240 km/h	Fast	Blue	1 yr	No
Car 4	200 km/h	Slow	Blue	5 yrs	Yes
Car 5	100 km/h	Fast	Yellow	5 yrs	Yes
Car 6	100 km/h	Slow	Black	6 yrs	No

Supervised Learning

Process of predicting values or categories

Pricing Example	Car Example	POS Example
€55.41 €80.12 €90.00 ⋮ €97.55	Accident Prone Not Accident Prone	Noun, Verb Pronoun, Adjective
Regression	Binary Classification	Multiclass Classification

Supervised Learning

Process of predicting values or categories

Pricing Example	Car Example	POS Example
€55.41 €80.12 €90.00 ⋮ €97.55	Accident Prone Not Accident Prone	Noun, Verb Pronoun, Adjective
Regression	Binary Classification	Multiclass Classification
Predicting continuous real values	Predicting two classes eg. yes or no	Predicting more than two classes

Binary Classification Exercise

Binary Classification Exercise

- We will start by looking at a simple technique - ***Linear* classification** for two classes

Binary Classification Exercise

- We will start by looking at a simple technique - ***Linear* classification** for two classes
- Our model/function will predict just one real number
- If this number is < 0, we will consider it to belong to **Class 1**. If it is ≥ 0, we will consider it to belong to **Class 2**.

Binary Classification Exercise

- Choose w_0, w_1 and b such that positive examples give a result > 0 and negative examples give a result < 0

$$w_0 \times x_0 + w_1 \times x_1 + b$$

x_0	x_1	Class
2	0	Positive
5	-2	Positive
-2	2	Negative
-1	-3	Negative

Binary Classification Exercise

- Choose w_0, w_1 and b such that positive examples give a result > 0 and negative examples gi

```
  0.3 x car.maxspeed
+ 0.2 x car.acceleration
+ 0.0 x car.color
+ 0.5 x car.age
```

$$w_0 \quad + b$$

x_0	x_1	
2	0	Positive
5	-2	Positive
-2	2	Negative
-1	-3	Negative

Binary Classification Exercise

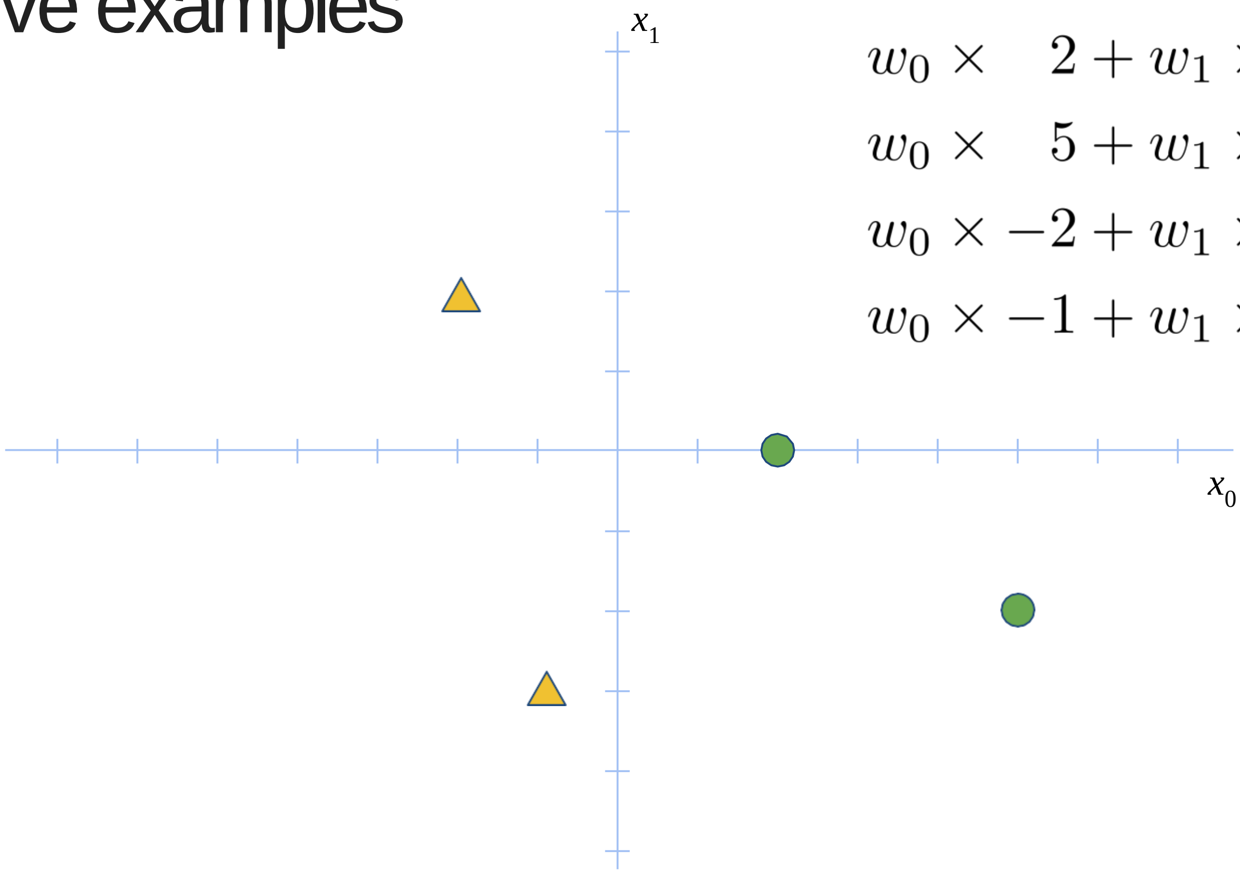
- Choose w_0, w_1 and b such that positive examples give a result > 0 and negative examples give a result < 0

$$w_0 \times x_0 + w_1 \times x_1 + b$$

x_0	x_1	**Class**
2	0	Positive
5	-2	Positive
-2	2	Negative
-1	-3	Negative

Binary Classification Exercise

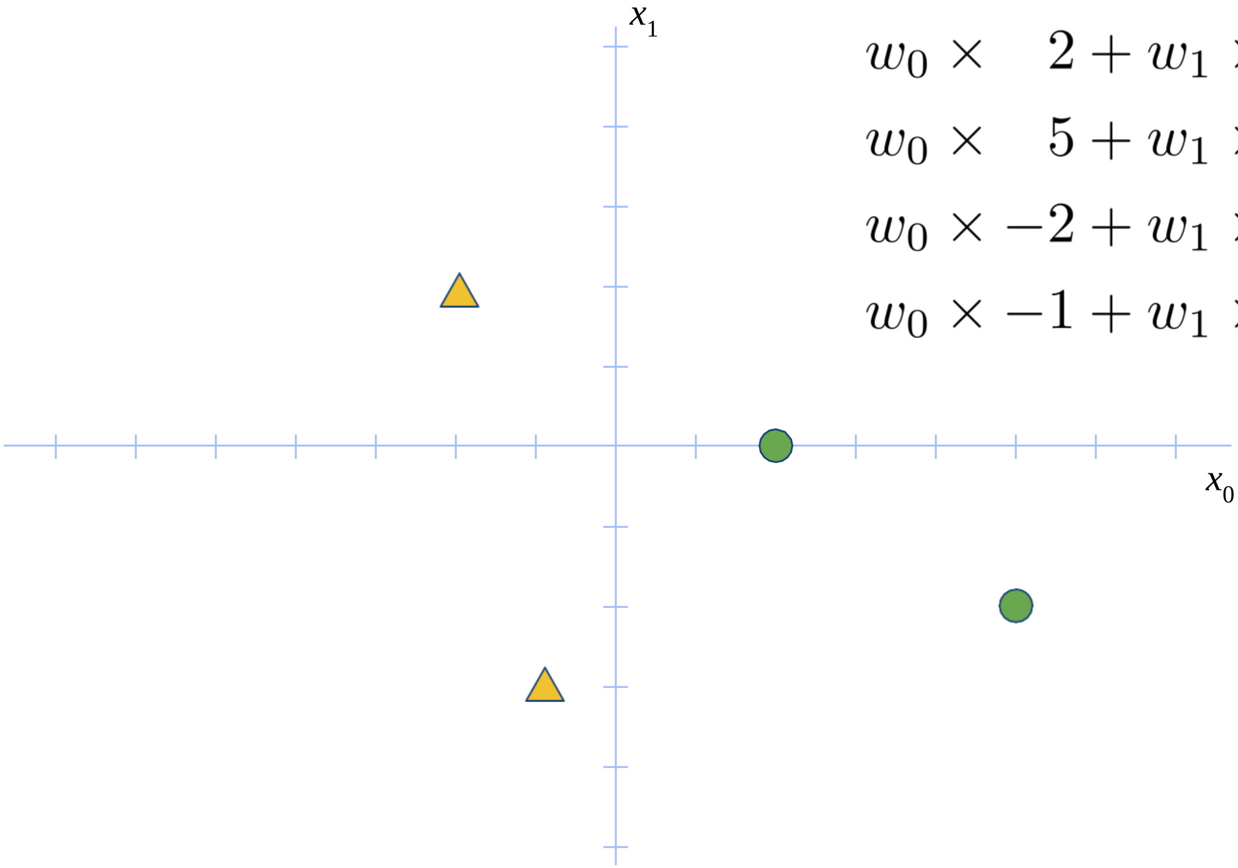
Find weights that separate positive examples from negative examples



$$w_0 \times 2 + w_1 \times 0 + b > 0$$
$$w_0 \times 5 + w_1 \times -2 + b > 0$$
$$w_0 \times -2 + w_1 \times 2 + b < 0$$
$$w_0 \times -1 + w_1 \times -3 + b < 0$$

Binary Classification Exercise

- Potential Solution: $w_0 = 3, w_1 = 1$ and $b = 3$



$$w_0 \times 2 + w_1 \times 0 + b > 0$$
$$w_0 \times 5 + w_1 \times -2 + b > 0$$
$$w_0 \times -2 + w_1 \times 2 + b < 0$$
$$w_0 \times -1 + w_1 \times -3 + b < 0$$

Binary Classification Exercise

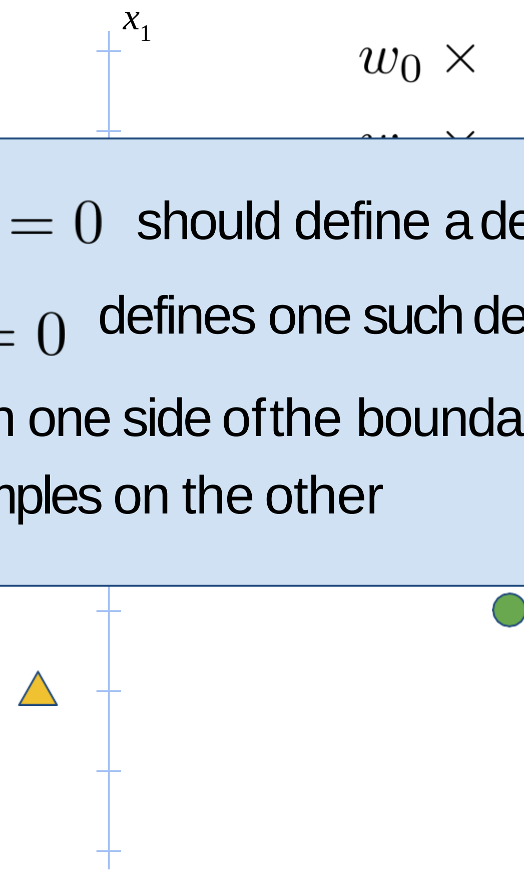
- Potential Solution: $w_0 = 3, w_1 = 1$ and $b = 3$

$$w_0 \times \quad 2 + w_1 \times \quad 0 + b > 0$$

$$w_0 \times x_0 + w_1 \times x_1 + b = 0$$ should define a decision boundary

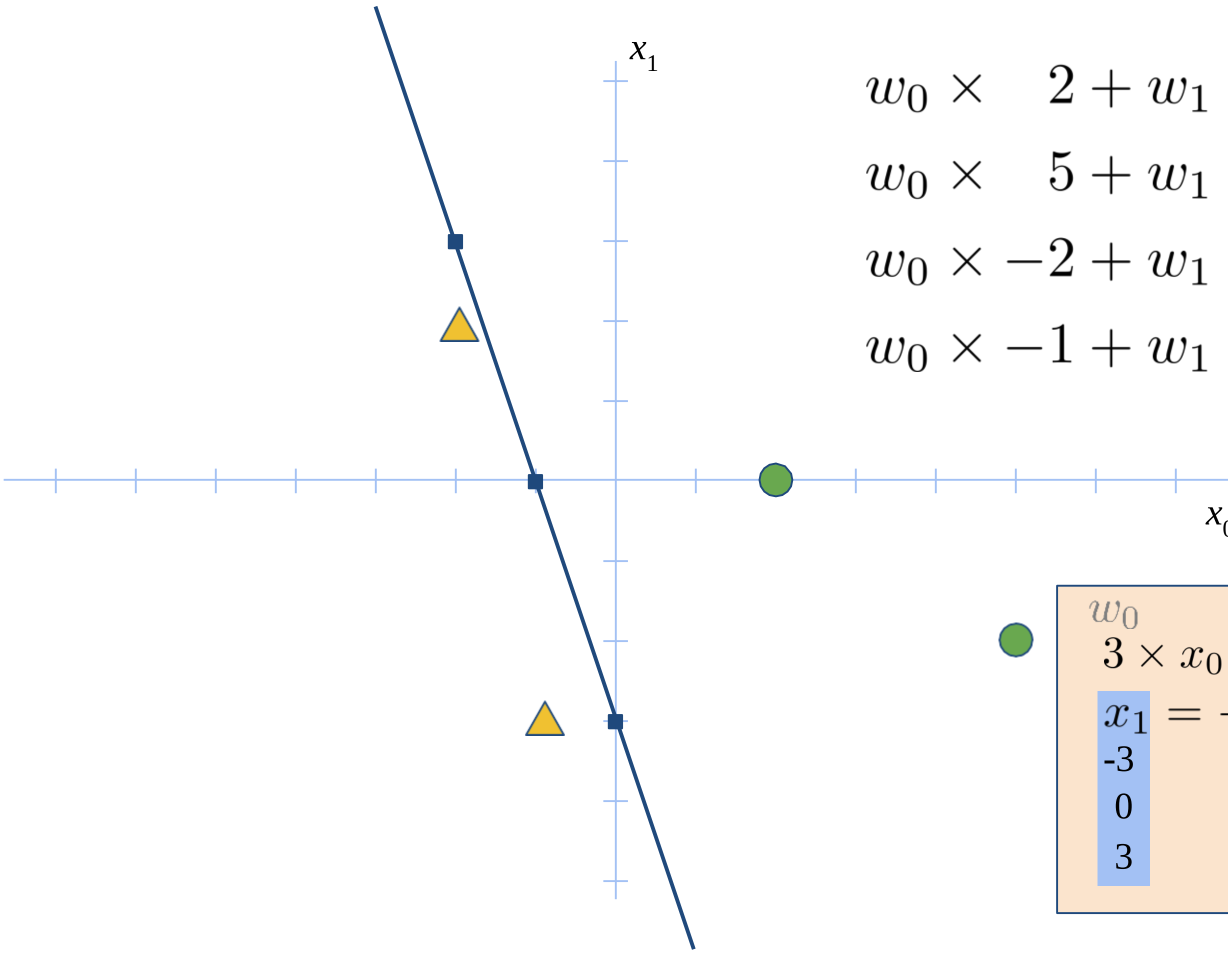
$$3 \times x_0 + 1 \times x_1 + 3 = 0$$ defines one such decision boundary

Positive examples will be on one side of the boundary, and negative examples on the other



Binary Classification Exercise

- Potential Solution: $w_0 = 3, w_1 = 1$ and $b = 3$



$$w_0 \times 2 + w_1 \times 0 + b > 0$$
$$w_0 \times 5 + w_1 \times -2 + b > 0$$
$$w_0 \times -2 + w_1 \times 2 + b < 0$$
$$w_0 \times -1 + w_1 \times -3 + b < 0$$

w_0 w_1 b

$$3 \times x_0 + 1 \times x_1 + 3 = 0$$
$$x_1 = -3 \times x_0 - 3$$

x_1	x_0
-3	0
0	-1
3	-2

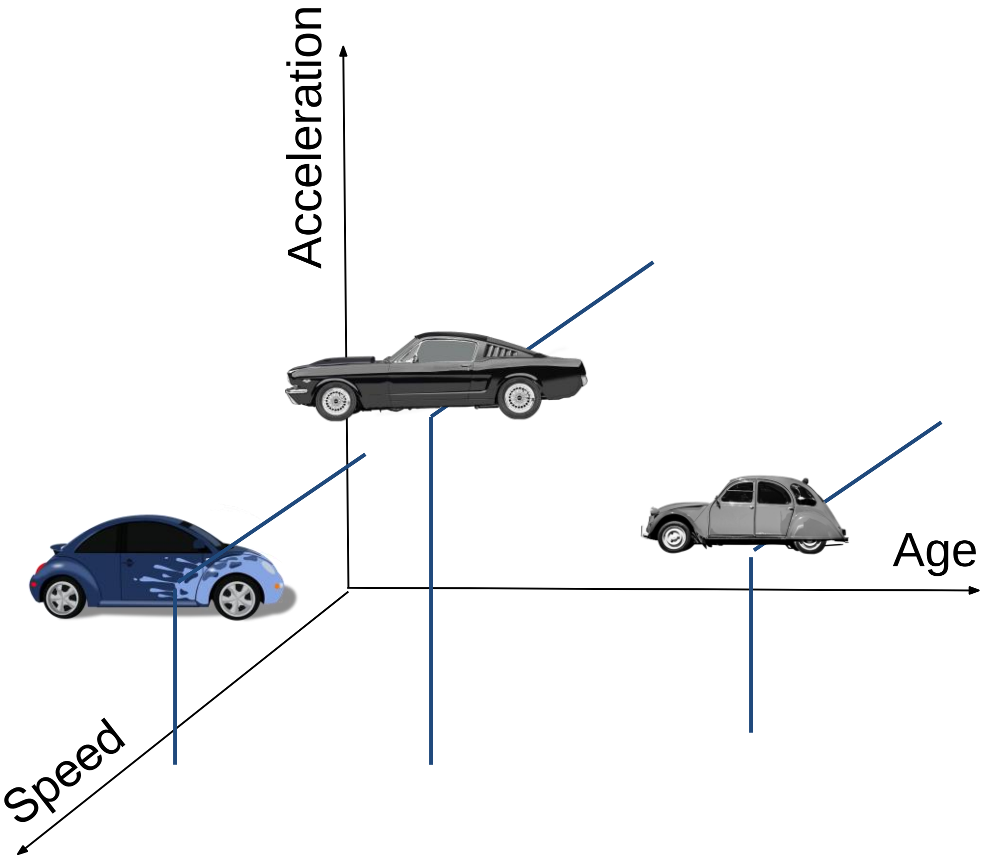
Binary Classification

What did we do?

- We were given a set of points in space
- We tried to draw a line to separate the "positive" points from the "negative" points
- The line was defined using "feature weights"

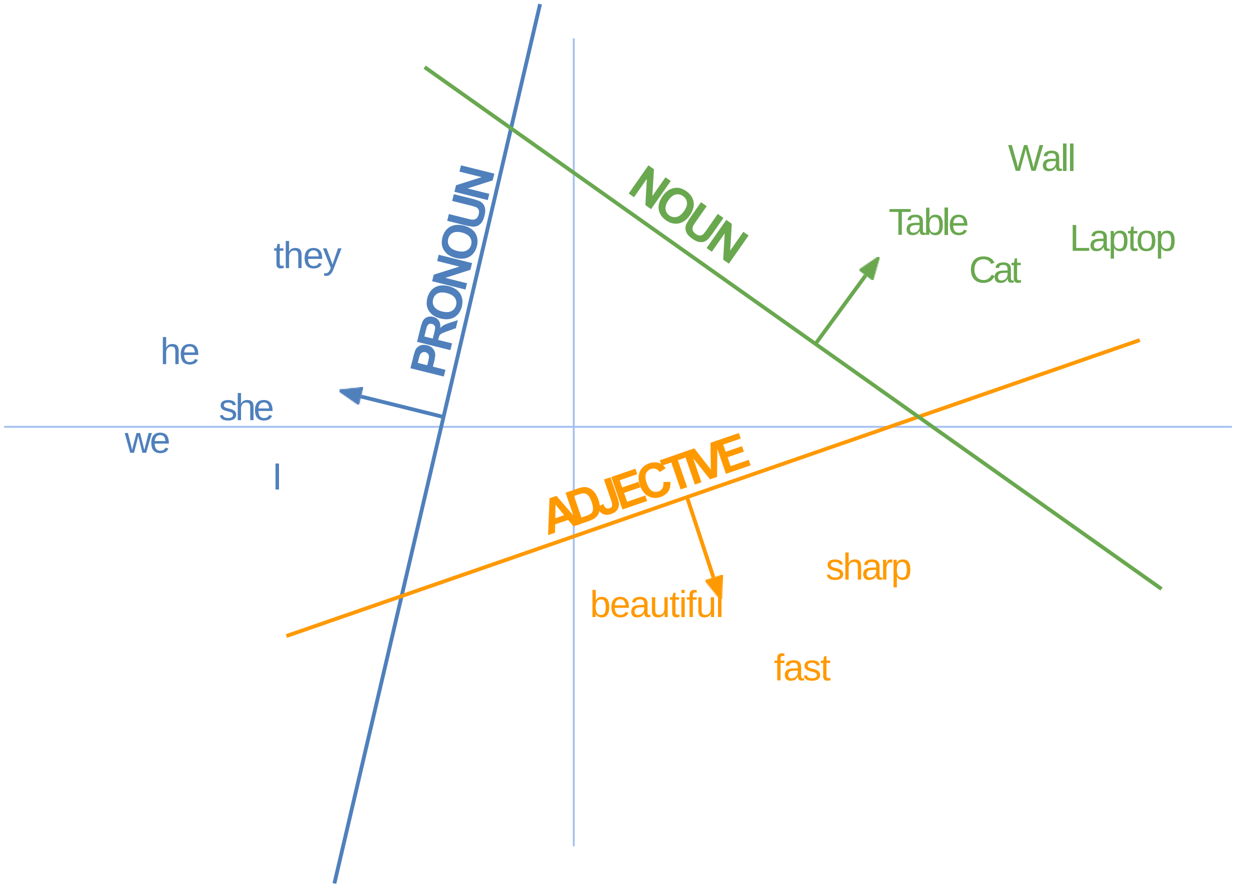
Vector Spaces

Vector Spaces



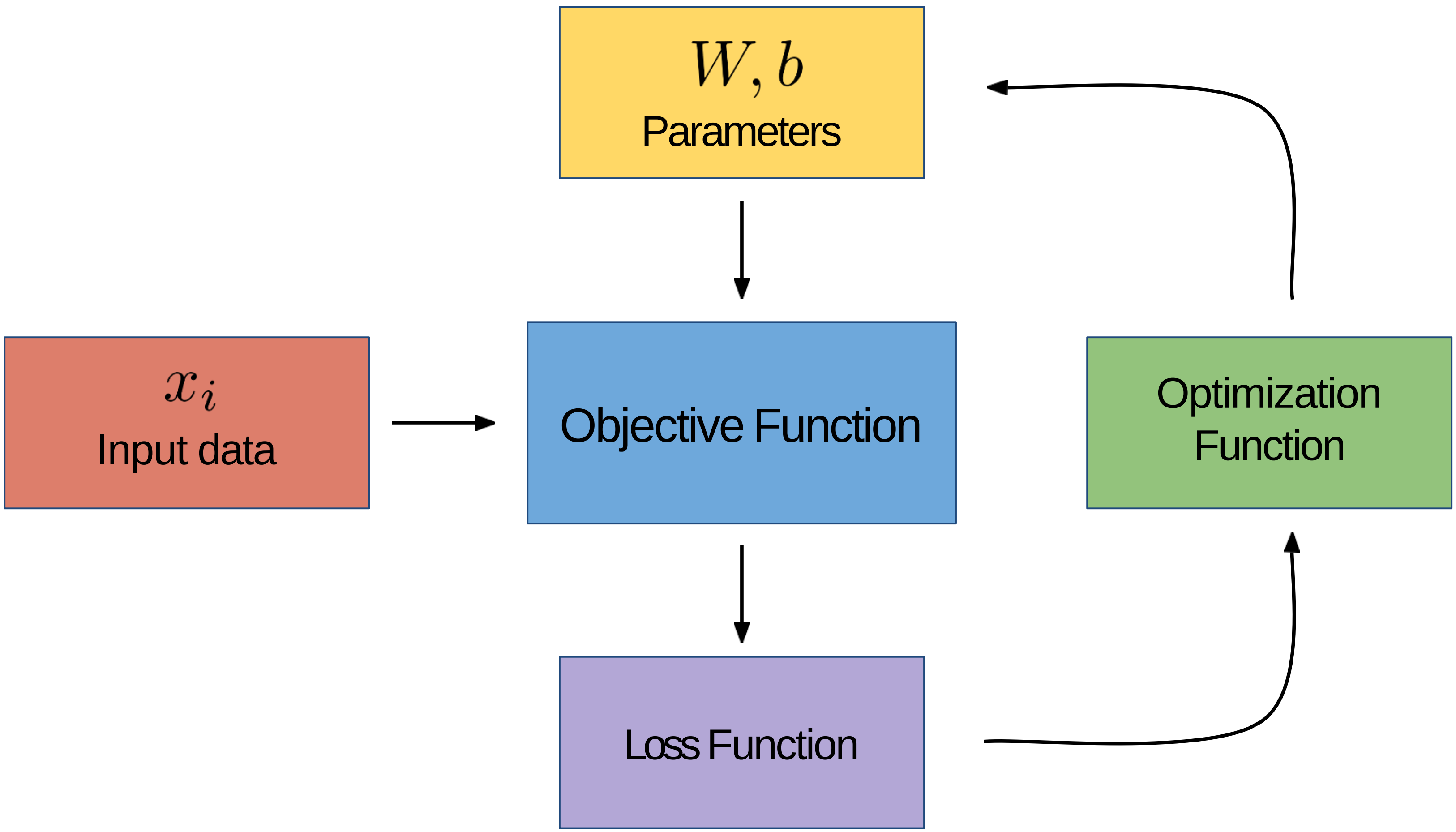
Imagine every feature as a dimension in space
Every object (car) can be represented as a point in space

Vector Spaces



Lifecycle of Training a Machine Learning Model

Learning Lifecycle



Objective Function

$$f(x, W, b) = W \cdot x + b$$

Objective function defines our *goal*

f ( , W, b) = N Numbers

Input

Parameters
P

Objective Function

$$f(x, W, b) = W \cdot x + b$$

Objective function defines our *goal*

f ( , W, b) = N Numbers

Input

Parameters
P

N can be 1 as in our previous example

Objective Function

f ( , W, b) = 1 Number

This 1 number can be a real valued output (for example depicting price, age etc). This is called regression.

This 1 number can also be used in the special case of binary classification (two classes) like we did in the previous exercise - i.e. **Class 1 if f > 0 and Class 2 if f ≤ 0**

Objective Function

$$f(x, W, b) = W \cdot x + b$$

Objective function defines our *goal*

f ( , W, b) = N Numbers

Input

Parameters
P

Outputs
N=2 can be used for binary classification
Example: scores for the two classes - *accident prone* and *not accident prone*

Objective Function

$$f(x, W, b) = W \cdot x + b$$

Objective function defines our *goal*

f ( , W, b) = N Numbers

Input

Parameters
P

Outputs
N=M can be used for M-class classification
Example: M = 40 for POS tags

Objective Function

$$f(x, W, b) = W \cdot x + b$$

Objective function defines our *goal*

f ( , W, b) = 2 Numbers

Input

Parameters
P

Outputs

Learned by the algorithm, just like we learned W and b in the previous exercise!

Learning Lifecycle