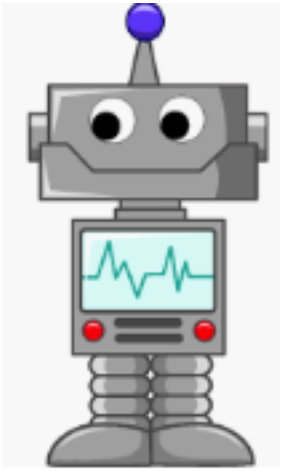
Lesson 2

My first machine learning model from Scratch

Teach a machine to identify vehicle types



Represent the sample

#Wheel	Height	Weight	Color

Represent the sample

#Wheel Height Weight Color

Identify the features which can represent the objects

$$F = \{f_1 f_2 f_3 \dots f_k\}$$

Feature set={ #Wheel Height Weight Color }

Represent the sample

#Wheel Height Weight Color

Identify the features which can represent the objects

$$F = \{f_1 f_2 f_3 \dots f_k\}$$

For every sample, assign value to corresponding feature

$$v_i = \{w_{i1} w_{i2} w_{i3} \dots w_{ik}\}$$

where w_{ij} is the value assigned for the feature f_j

Represent the sample

#Wheel	Height	Weight	Color
4	6	500	Red
4	5.5	600	Blue
4	5	550	Yellow
2	3	200	Red
2	3.5	150	blue
2	4	250	Yellow

For every object, assign value to corresponding feature

$$v_i = \{w_{i1} w_{i2} w_{i3} \dots w_{ik}\}$$

where w_{ij} is the value assigned for the feature f_j

Vector Space Model

#Wheel	Height	Weight	Color
4	6	500	Red
4	5.5	600	Blue
4	5	550	Yellow
2	3	200	Red
2	3.5	150	blue
2	4	250	Yellow

Features Vectors

This form of representation is called Vector Space Model

Are all features useful?

#Wheel	Height	Weight	Color
4	6	500	Red
4	5.5	600	Blue
4	5	550	Yellow
2	3	200	Red
2	3.5	150	blue
2	4	250	Yellow

Features

Features Vectors

Good Features

- #Wheel
- Height
- Weight

Bad Feature

- Colour

Let us consider single feature

#Wheel	Class Label
4	CAR
4	CAR
4	CAR
2	BIKE
2	BIKE
2	BIKE

Training Dataset

Feature vector with Class label

Given the #Wheel, identify the vehicle

	#Wheel	Class Label
	4	CAR
	4	CAR
	4	CAR
	2	BIKE
	2	BIKE
	2	BIKE
	2	?

Let us estimate

#Wheel	Class Label
4	CAR
4	CAR
4	CAR
2	BIKE
2	BIKE
2	BIKE

Pr(Vehicle type| #Wheel) = ?

Let us estimate the probability (type|#wheel)

#Wheel	Class Label
4	CAR
4	CAR
4	CAR
2	BIKE
2	BIKE
2	BIKE

Pr(CAR| 4) = 100%

Pr(BIKE| 4) = 0%

Pr(CAR| 2) = 0%

Pr(BIKE| 2) = 100%

Ask the question now

	#Wheel	Class Label
	4	CAR
	4	CAR
	4	CAR
	2	BIKE
	2	BIKE
	2	BIKE
	2	

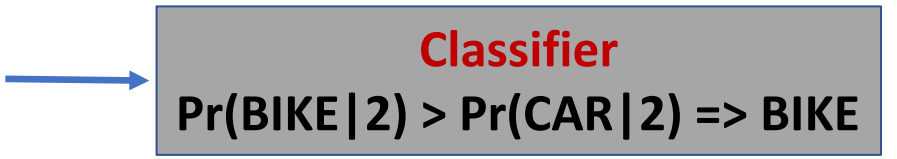
Pr(CAR| 4) = 100%

Pr(BIKE| 4) = 0%

Pr(CAR| 2) = 0%

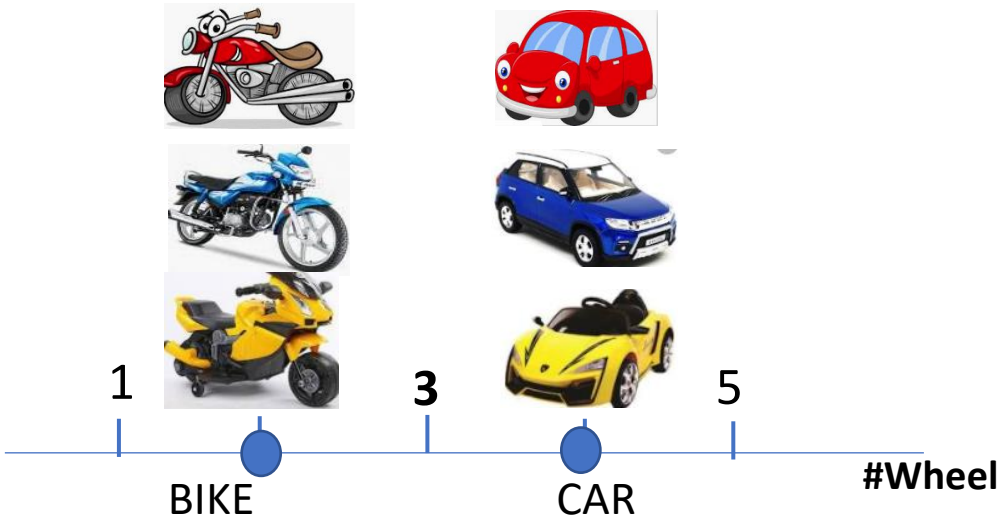
Pr(BIKE| 2) = 100%

{2} ?



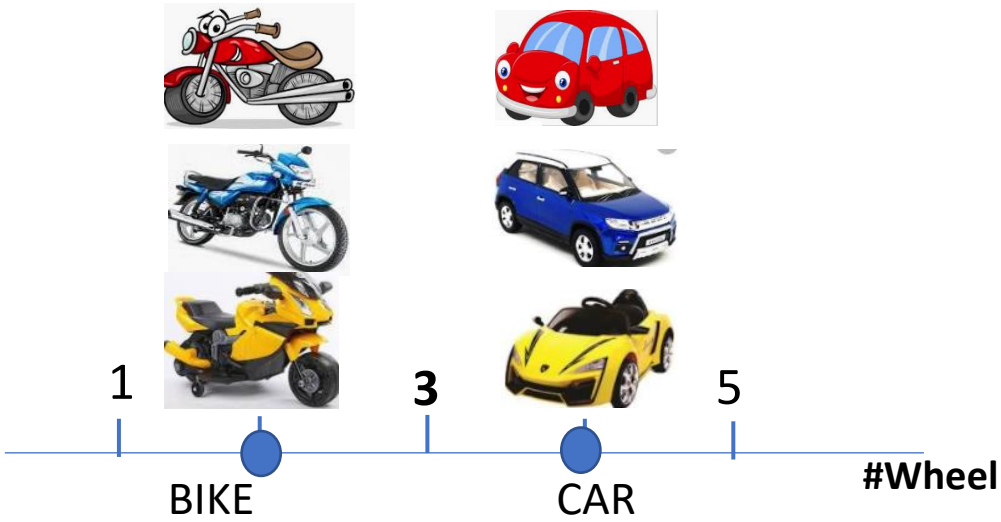
There are multiple ways

#Wheel	Class Label
4	CAR
4	CAR
4	CAR
2	BIKE
2	BIKE
2	BIKE



There are multiple ways

	#Wheel	Class Label
	4	CAR
	4	CAR
	4	CAR
	2	BIKE
	2	BIKE
	2	BIKE
	2	



2 → **Classifier**
If #Wheel < 3, then it is BIKE → **BIKE**

If selected feature is not sufficient

	#Wheel	Class Label
	4	CAR
	4	CAR
	4	CAR
	2	BIKE
	2	BIKE
	2	BIKE
	4	BIKE
	2	CAR

Pr(CAR| 4) = 75%

Pr(BIKE| 4) = 25%

Pr(CAR| 2) = 25%

Pr(BIKE| 2) = 75%

2 ?

If selected feature is not sufficient

	#Wheel	Class Label
	4	CAR
	4	CAR
	4	CAR
	2	BIKE
	2	BIKE
	2	BIKE
	4	BIKE
	2	CAR

Pr(CAR| 4) = 75%

Pr(BIKE| 4) = 25%

Pr(CAR| 2) = 25%

Pr(BIKE| 2) = 75%

2 **BIKE**

Pr(BIKE|2) > Pr(CAR|2) => BIKE

More Features

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

H: High, height >= 5

L: Low, height < 5

Estimate the probabilities, and ask the same question

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

Pr(CAR| 2,L) = 0%

Pr(BIKE| 2,H) = 0%

 {2 H} ?

Estimate the probabilities, and ask the same question

	#Wheel	Height	Class Label
	4	H	CAR
	4	H	CAR
	4	H	CAR
	2	L	BIKE
	2	L	BIKE
	2	L	BIKE
	4	L	BIKE
	2	H	CAR

Pr(CAR| 4,H) = 100%

Pr(BIKE| 4,L) = 100%

Pr(CAR| 2,H) = 100%

Pr(BIKE| 2,L) = 100%

Pr(CAR| 4,L) = 0%

Pr(BIKE|4,H) = 0%

Pr(CAR| 2,L) = 0%

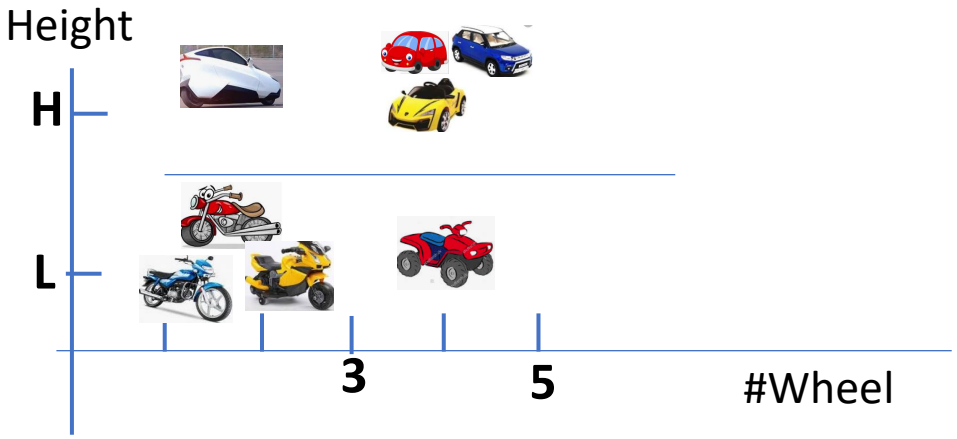
Pr(BIKE| 2,H) = 0%

{2 H}

CAR

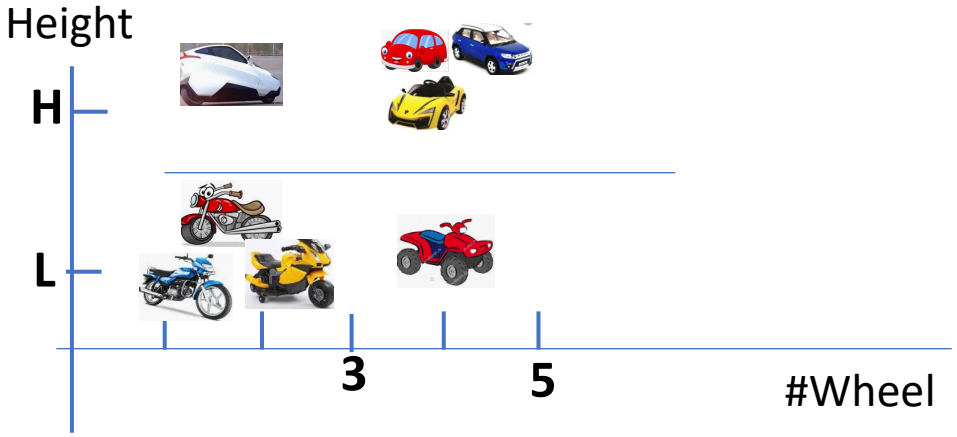
Multiple ways

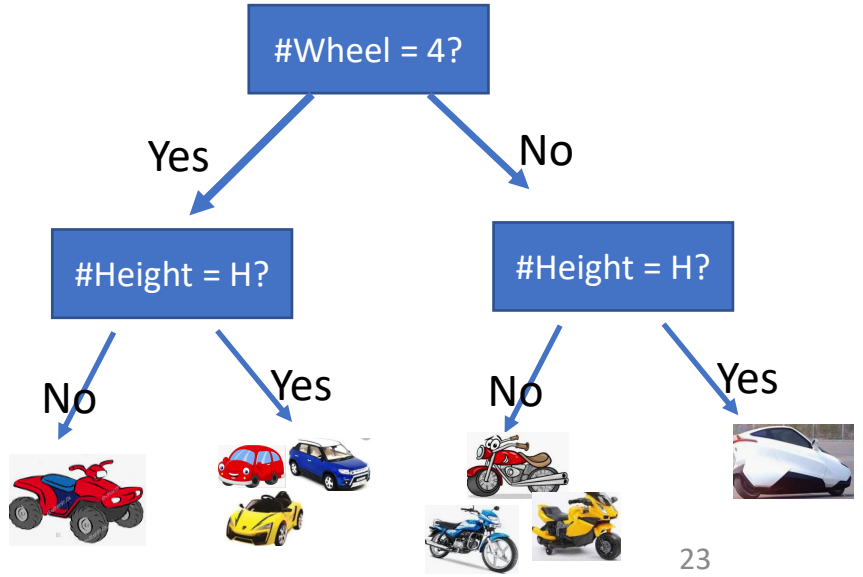
#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR



Multiple ways

#Wheel	Height	Class Label
4	H	CAR
4	H	CAR
4	H	CAR
2	L	BIKE
2	L	BIKE
2	L	BIKE
4	L	BIKE
2	H	CAR





Summary

- **Identify the features**
- **Represent the vehicles by the features**
- **Remove non-informative features**
- **Build the classification model from the data**
- **Perform the classification task**