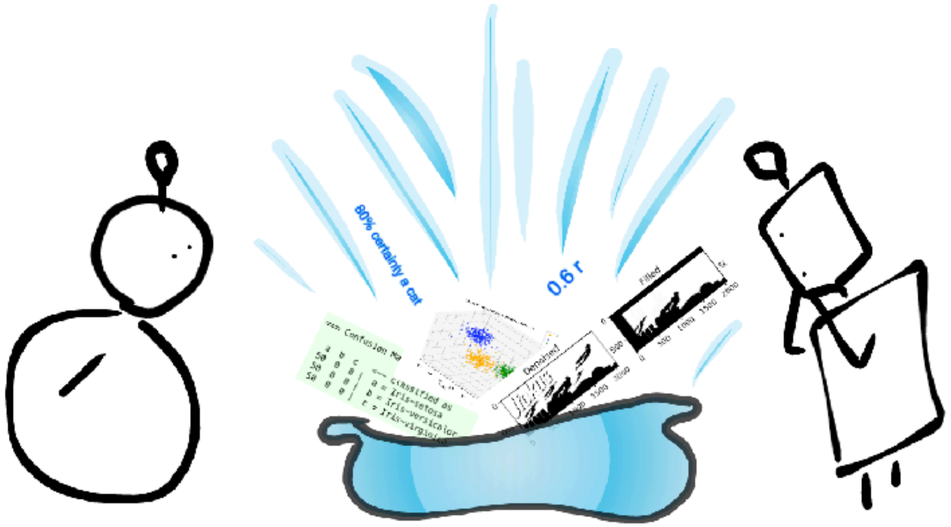
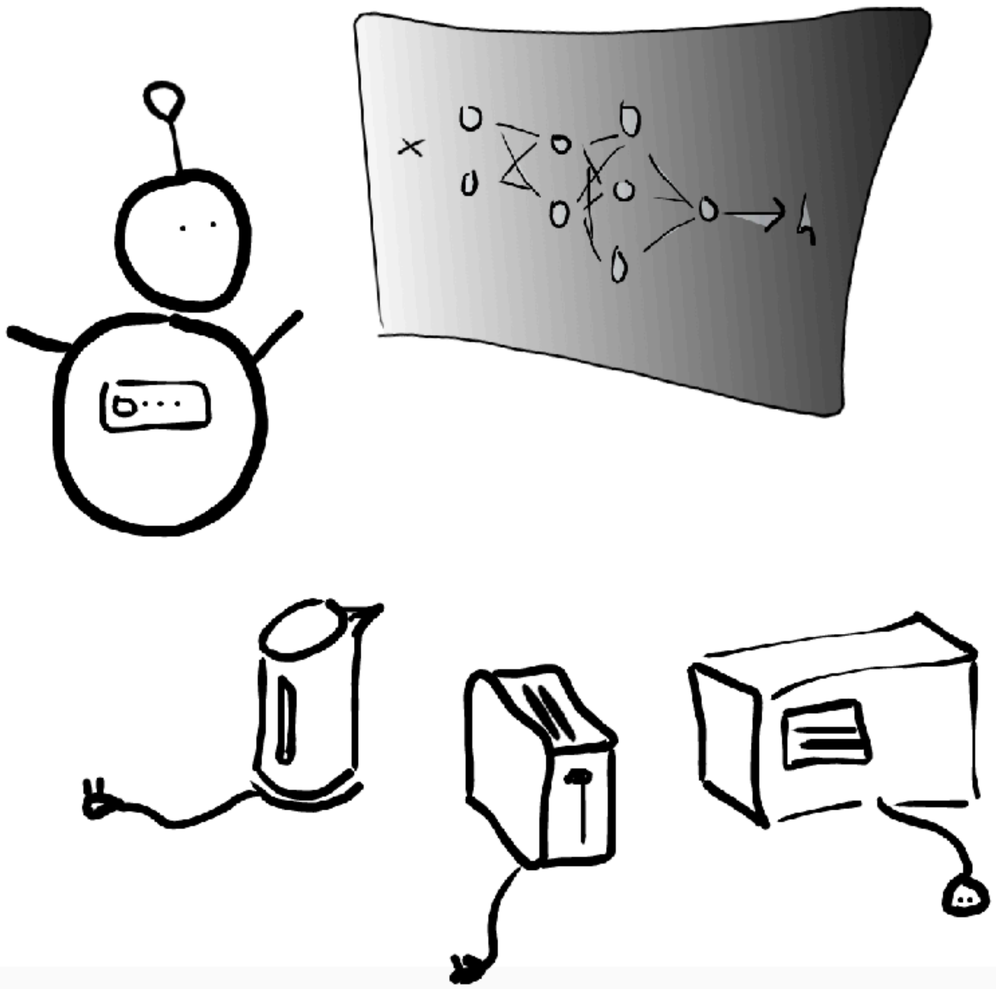
Machine Learning Bag o' Tricks



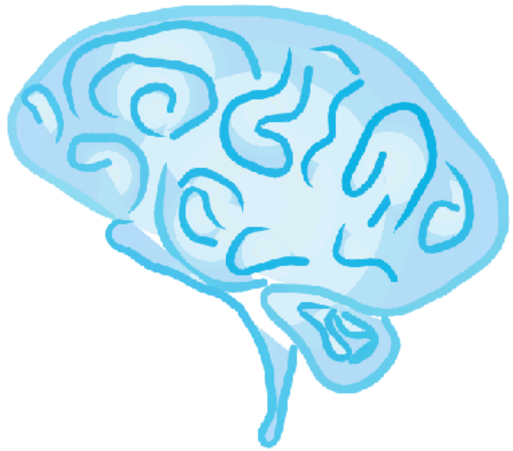
@katharinecodes



@katharinecodes

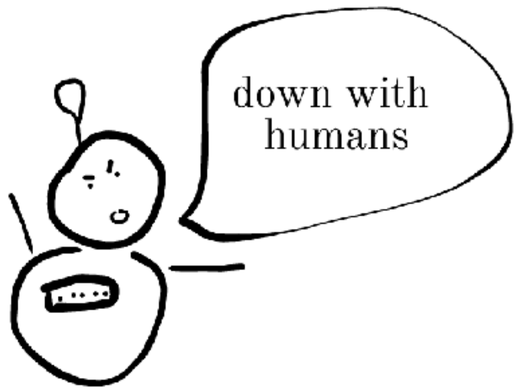
computers learning without being
explicitly programmed

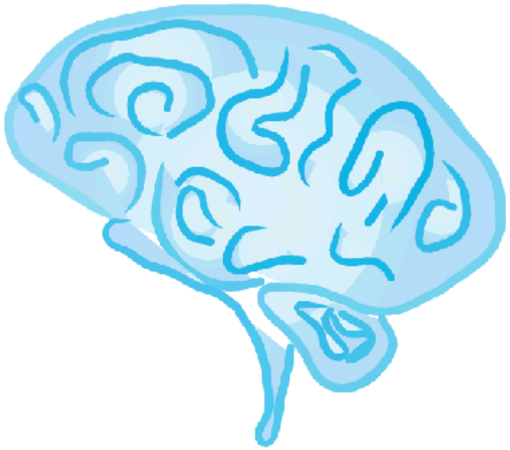
@katharinecodes



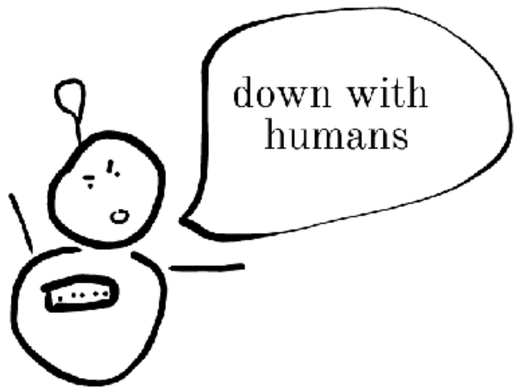
@katharinecodes

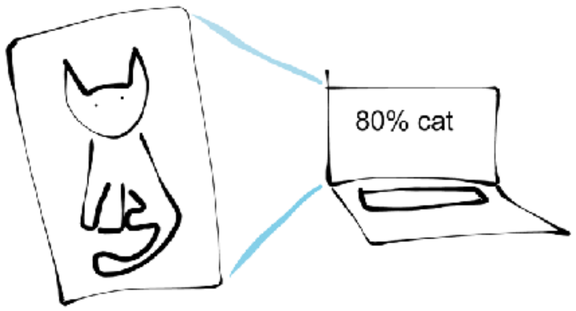
@katharinecodes



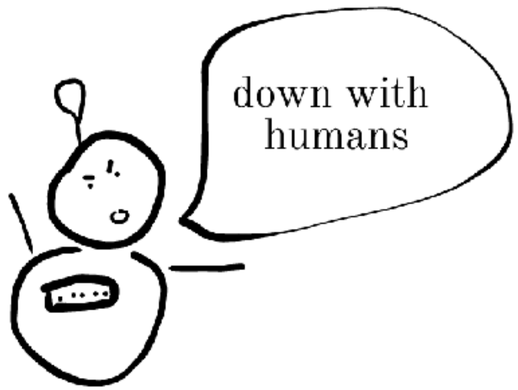


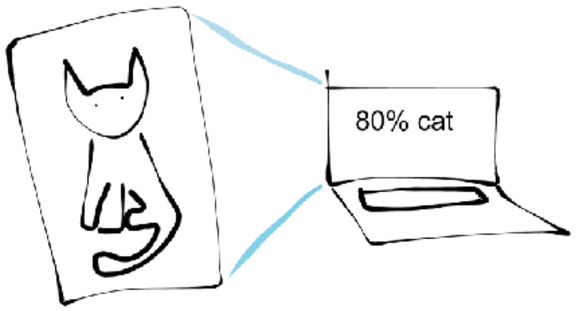
@katharinecodes

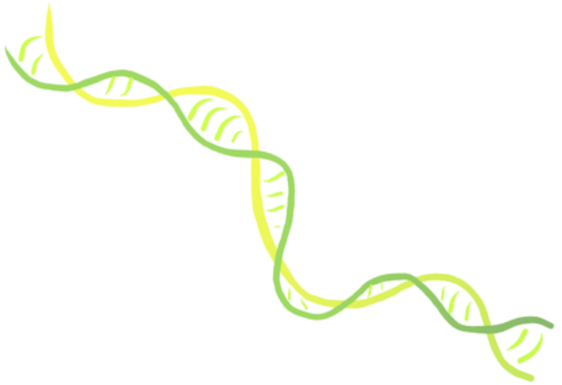




@katharinecodes





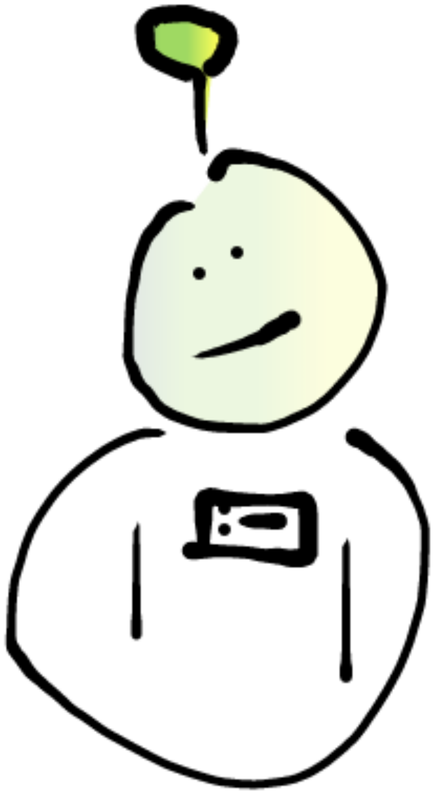


@katharinecodes

!HEALTH WARNING!

@katharinecodes

!HEALTH WARNING!



@katharinecodes

Your next 50 minutes…

@katharinecodes

Your next 50 minutes…

- Can I apply Machine Learning techniques to my data?

Your next 50 minutes…

- Can I apply Machine Learning techniques to my data?
- Can I find patterns, without knowing in advance what the patterns might be?

@katharinecodes

Your next 50 minutes…

- Can I apply Machine Learning techniques to my data?
- Can I find patterns, without knowing in advance what the patterns might be?
- How can I categorise things?

@katharinecodes

Your next 50 minutes…

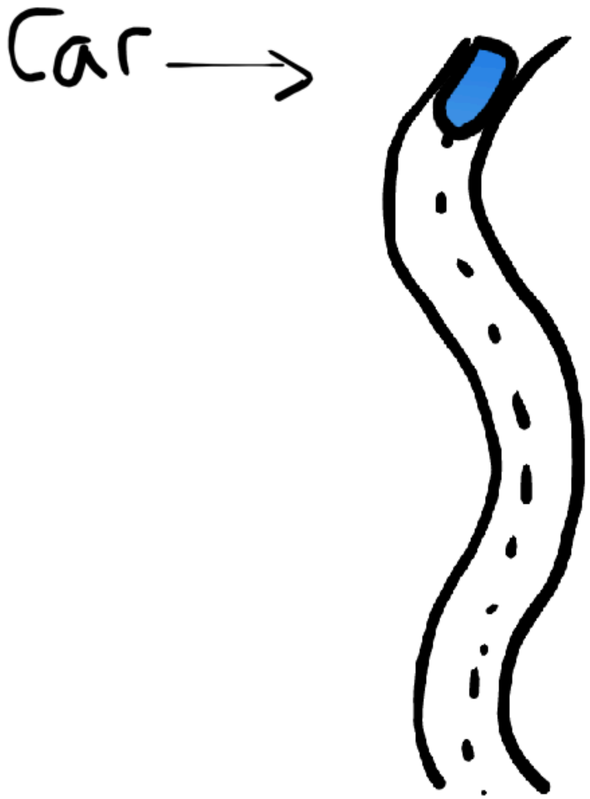
- Can I apply Machine Learning techniques to my data?
- Can I find patterns, without knowing in advance what the patterns might be?
- How can I categorise things?
- Can I predict values?

@katharinecodes

Roadmap

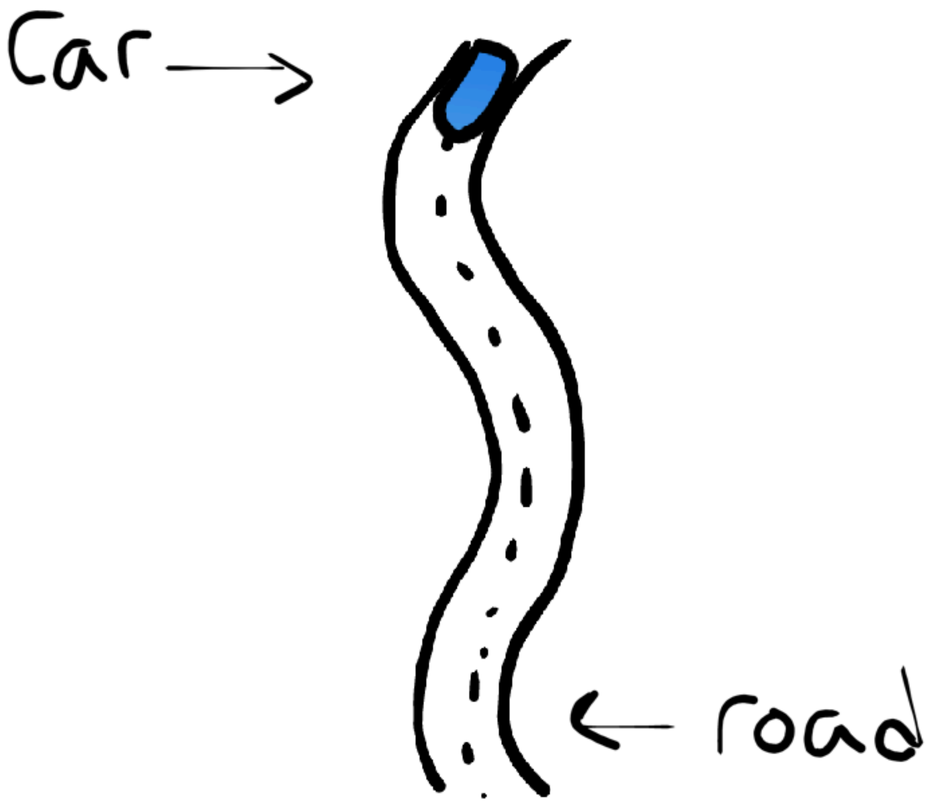
- Data
- Clustering
- Classification
- Regression

Roadmap



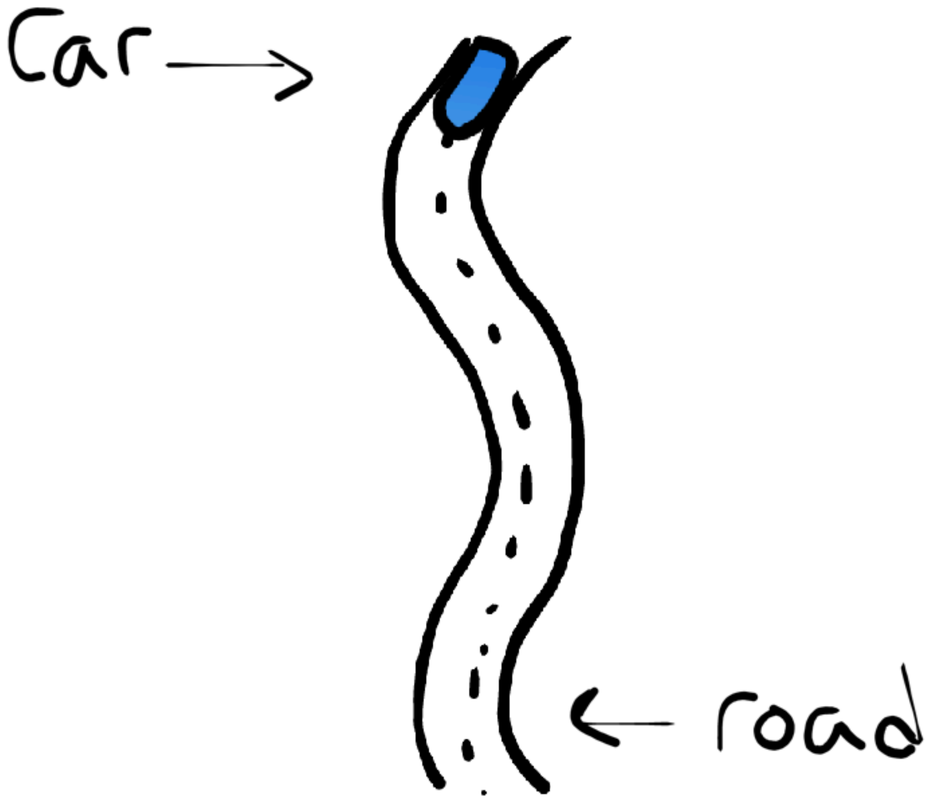
- Data
- Clustering
- Classification
- Regression

Roadmap

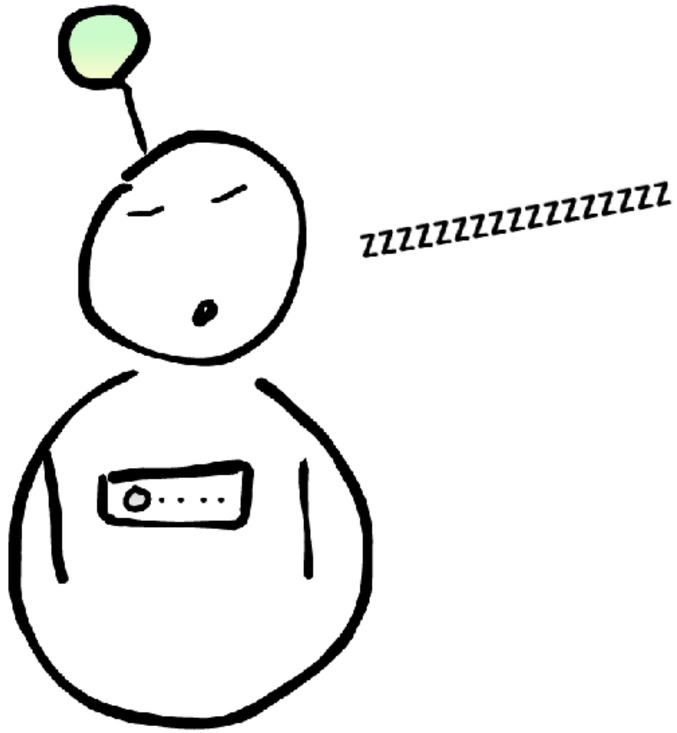


- Data
- Clustering
- Classification
- Regression

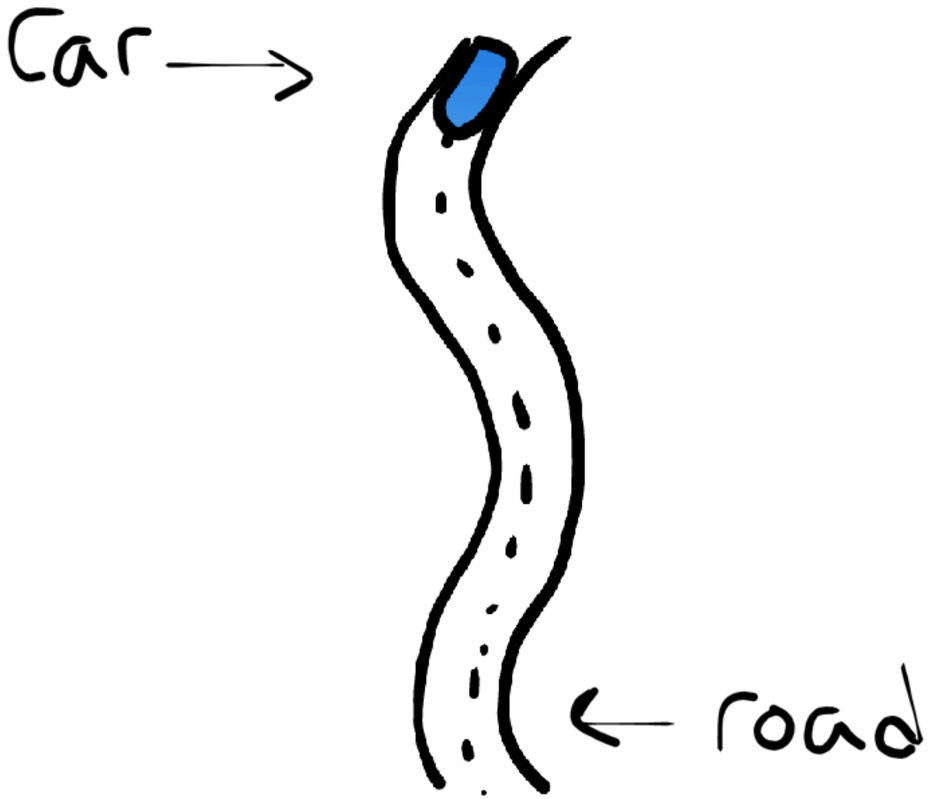
Roadmap



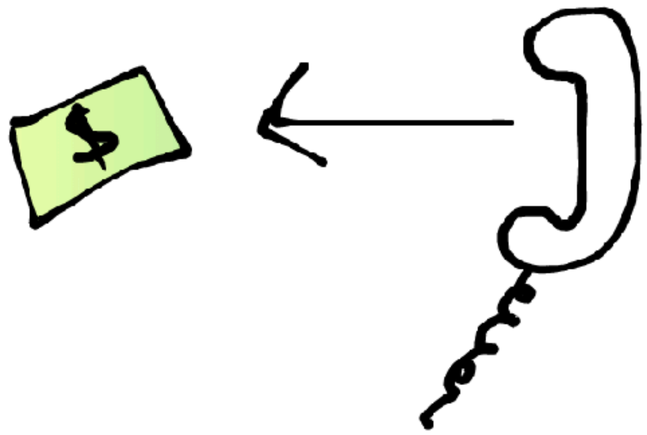
- Data
- Clustering
- Classification
- Regression


zzzzzzzzzzzzzzzz

Roadmap



- Data
- Clustering
- Classification
- Regression



What are you trying to work out?
[What's your hypothesis]

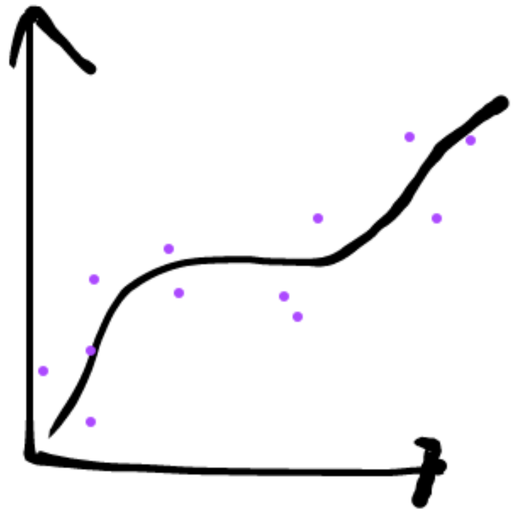
What are you trying to work out?
[What's your hypothesis]

Does your data lend itself to the hypothesis?

Regression

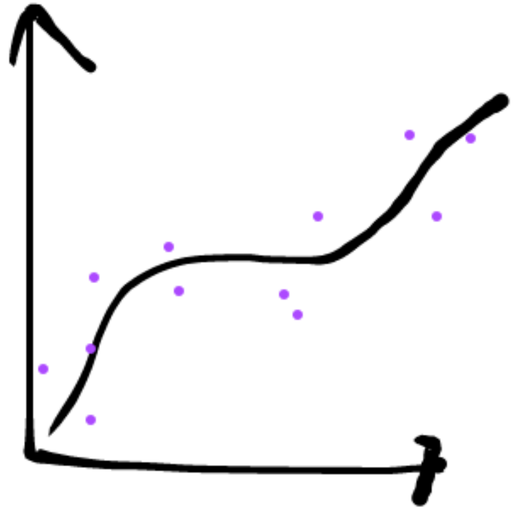
Classification

Regression



Classification

Regression



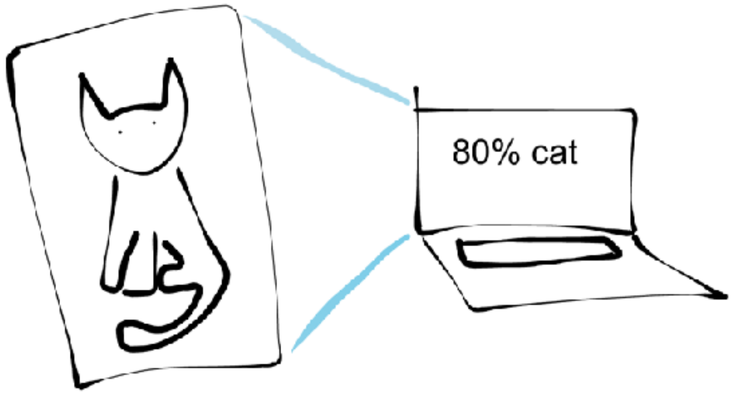
Classification

or

Supervised

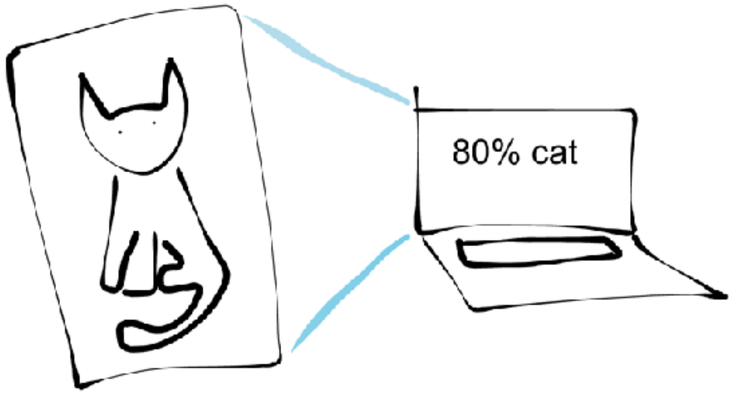
Unsupervised

Supervised

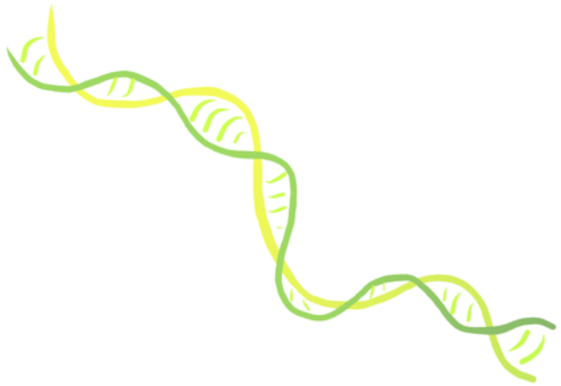


Unsupervised

Supervised

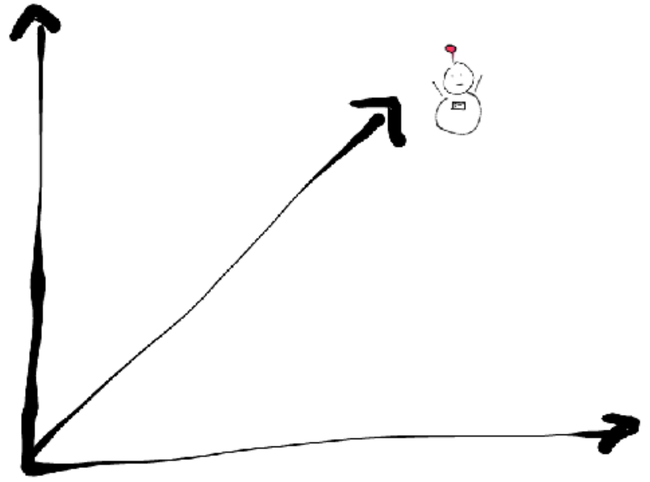


Unsupervised

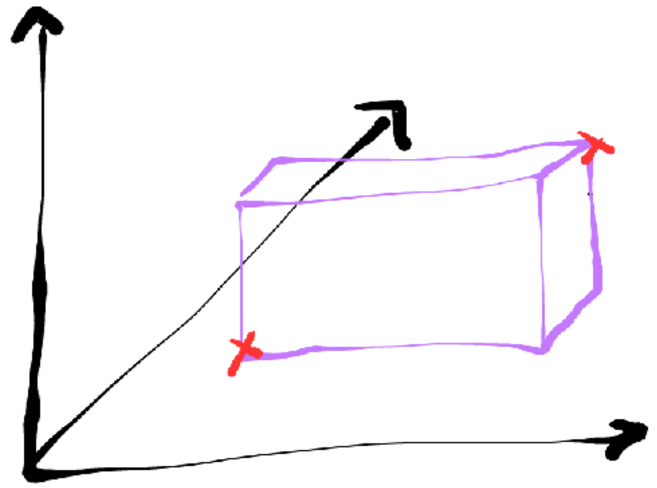


What is a feature?

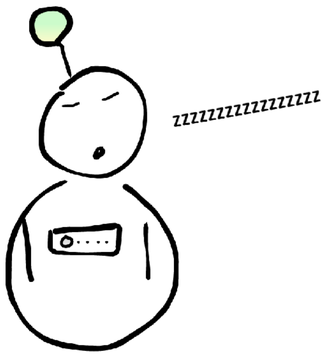
The Curse of Dimensionality



The Curse of Dimensionality

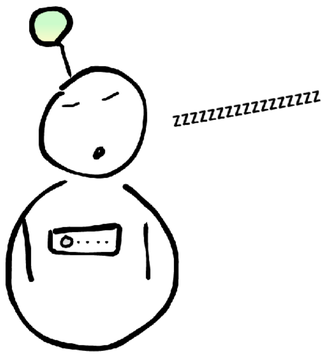


Data…



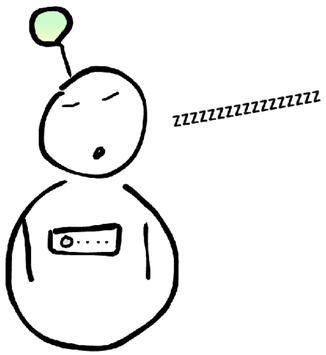
Data…

- Filters



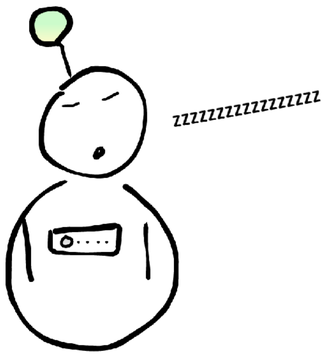
Data…

- Filters
- Wrappers



Data…

- Filters
- Wrappers
- PCA



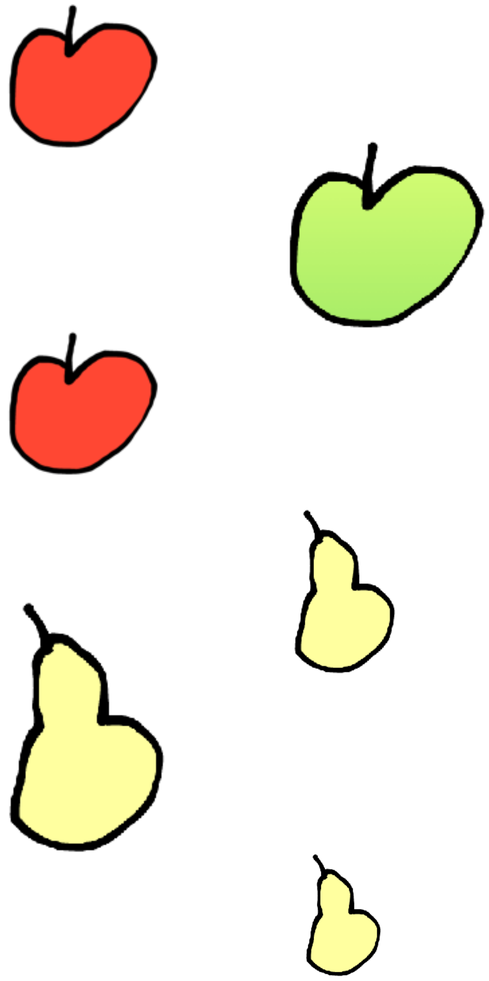
Features

Alcohol
Malic acid
Ash
Alcalinity of ash
Magnesium
Total phenols
Flavanoids
Non-flavanoid phenols
Proanthocyanins
Colour intensity
Hue
Proline
OD280/OD315 of diluted wines

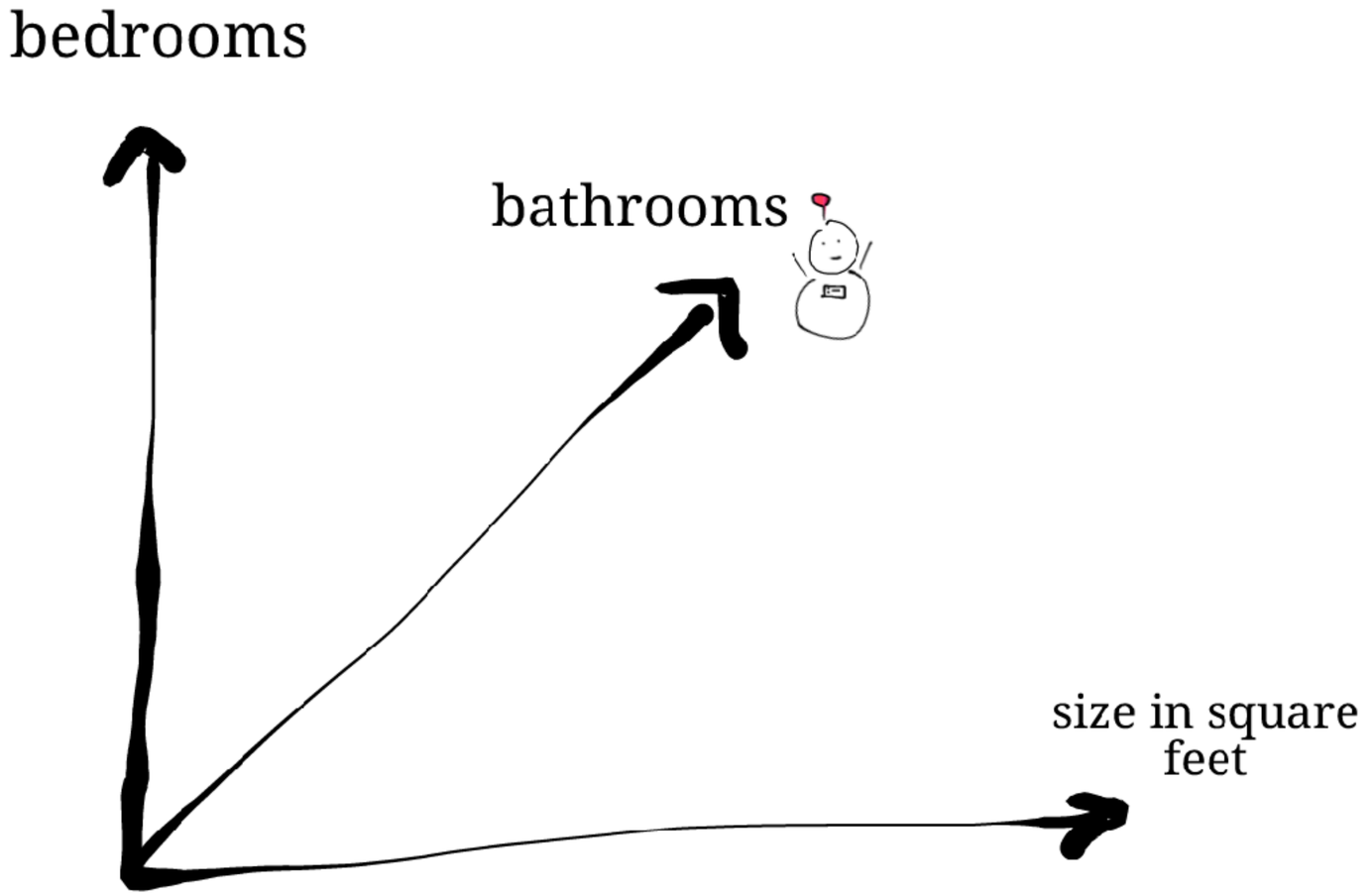
Class labels

Type 1
Type 2
Type 3

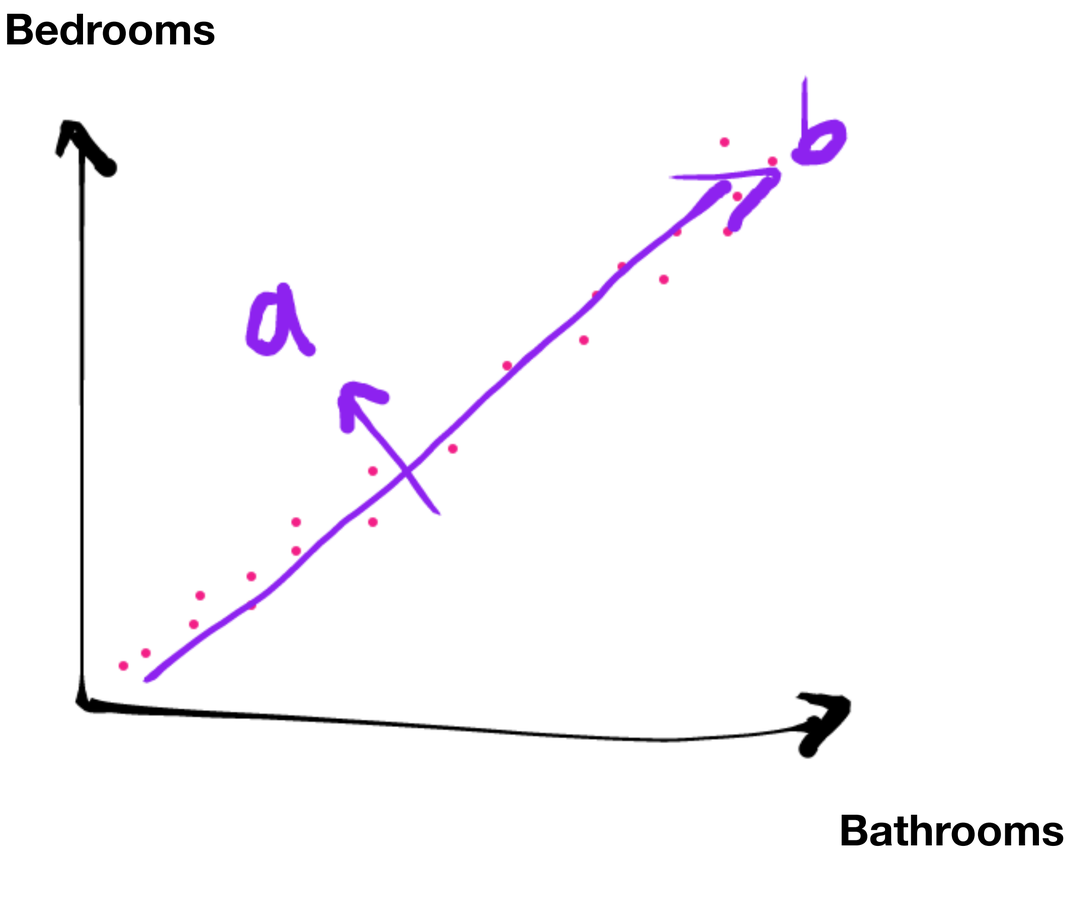
Features		Class label
Size	Colour	Fruit
Medium	Red	Apple
Large	Green	Apple
Medium	Red	Apple
Medium	Yellow	Pear
Large	Yellow	Pear
Small	Yellow	Pear

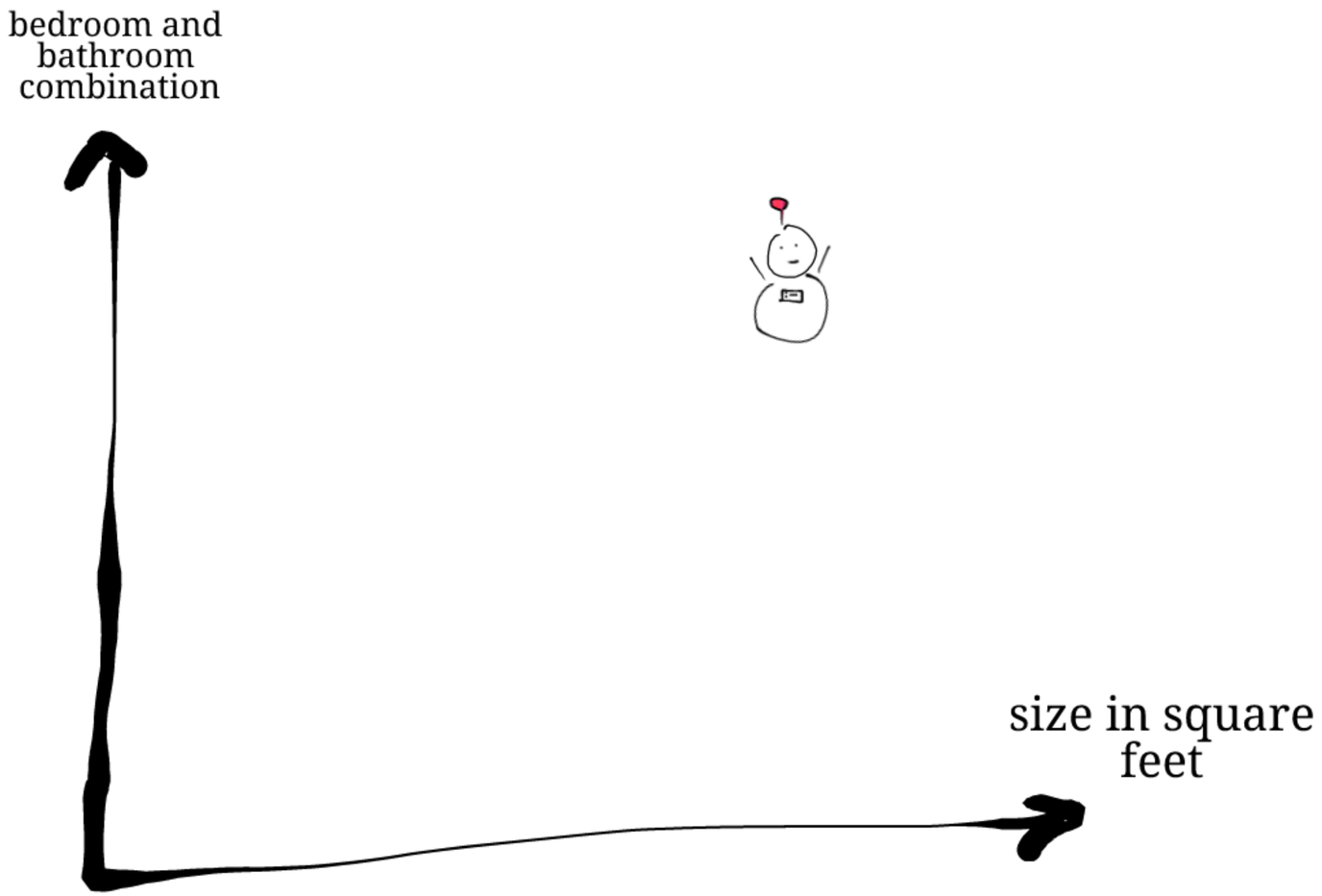


Filter	Wrapper
Flavanoids	Flavanoids
Proline	Hue
Colour intensity	Proline
OD280/OD315 of diluted wines	Alcohol
Alcohol	Colour intensity
Hue	Ash


bedrooms
bathrooms
size in square feet


Bedrooms
Bathrooms


Bedrooms
a
b
Bathrooms


bedroom and
bathroom
combination
size in square
feet

Roadmap

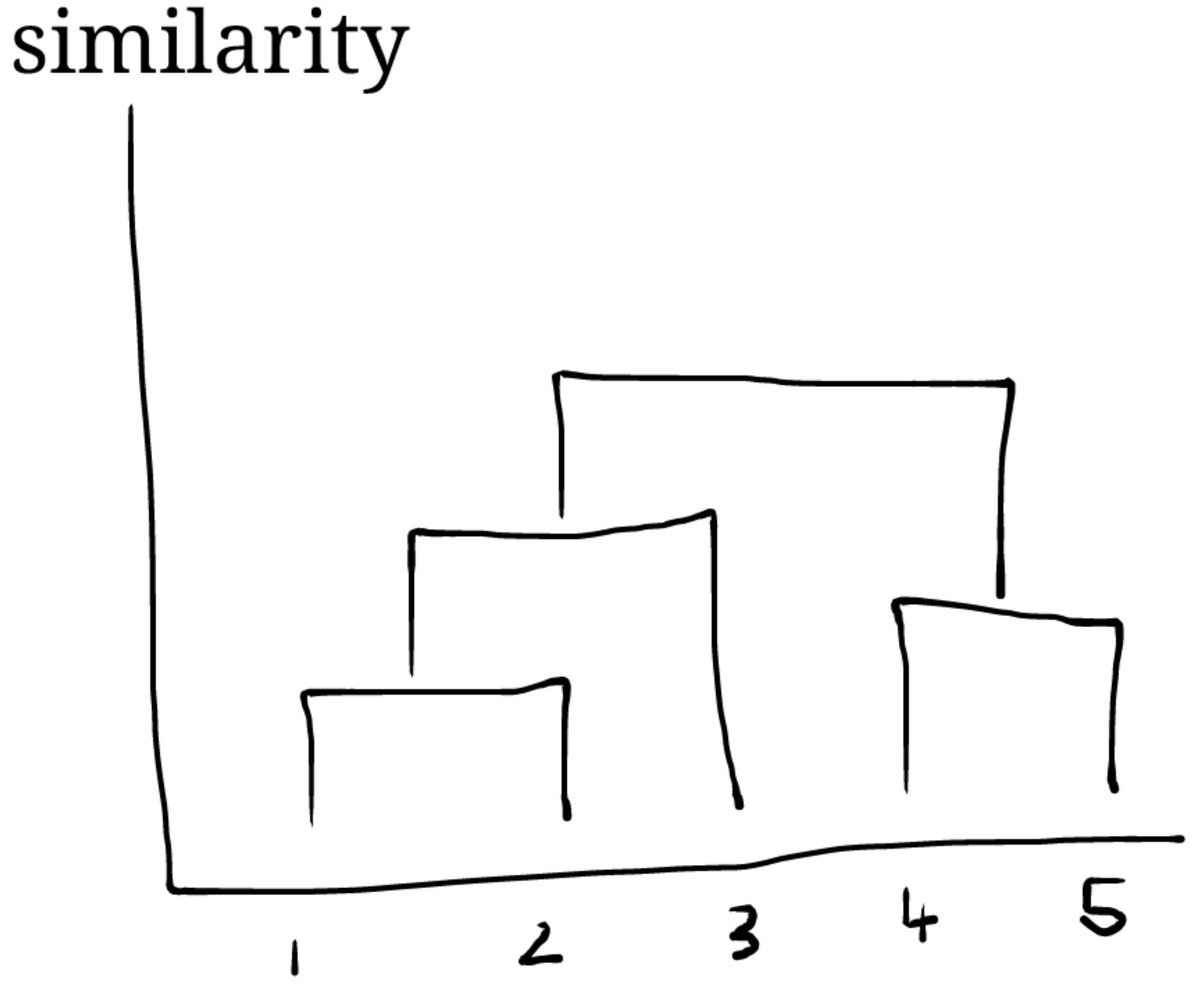
- Data
- Clustering
- Classification
- Regression

K Means Clustering

- Pick “k” cluster centroids
- Repeat until convergence:
  - For each data point, which centroid is it closest to? Assign it to that centroid.
  - Work out the centre of all of the data points assigned to the centroid. Move the centroid to that location.

Agglomerative clustering

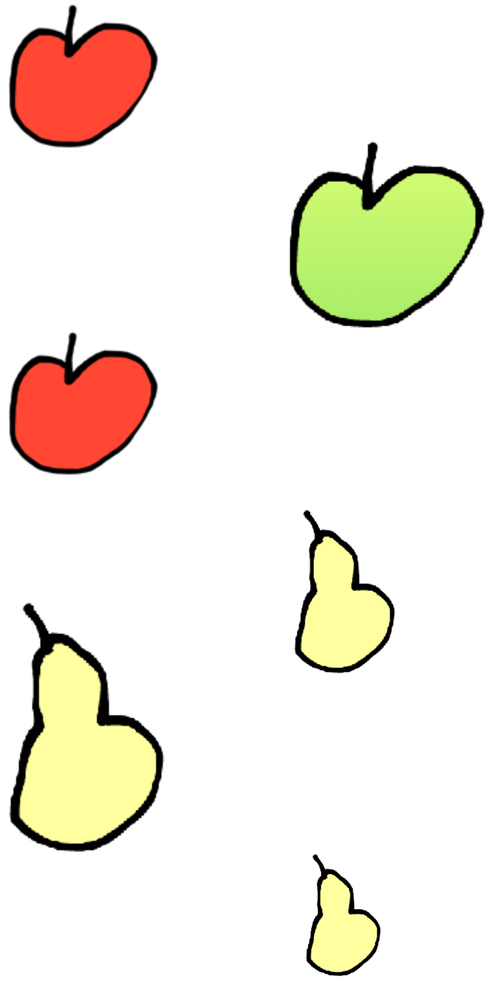
- Assign each item to it’s own cluster
- At each step, merge the most similar pairs of clusters
- Repeat until everything is merged

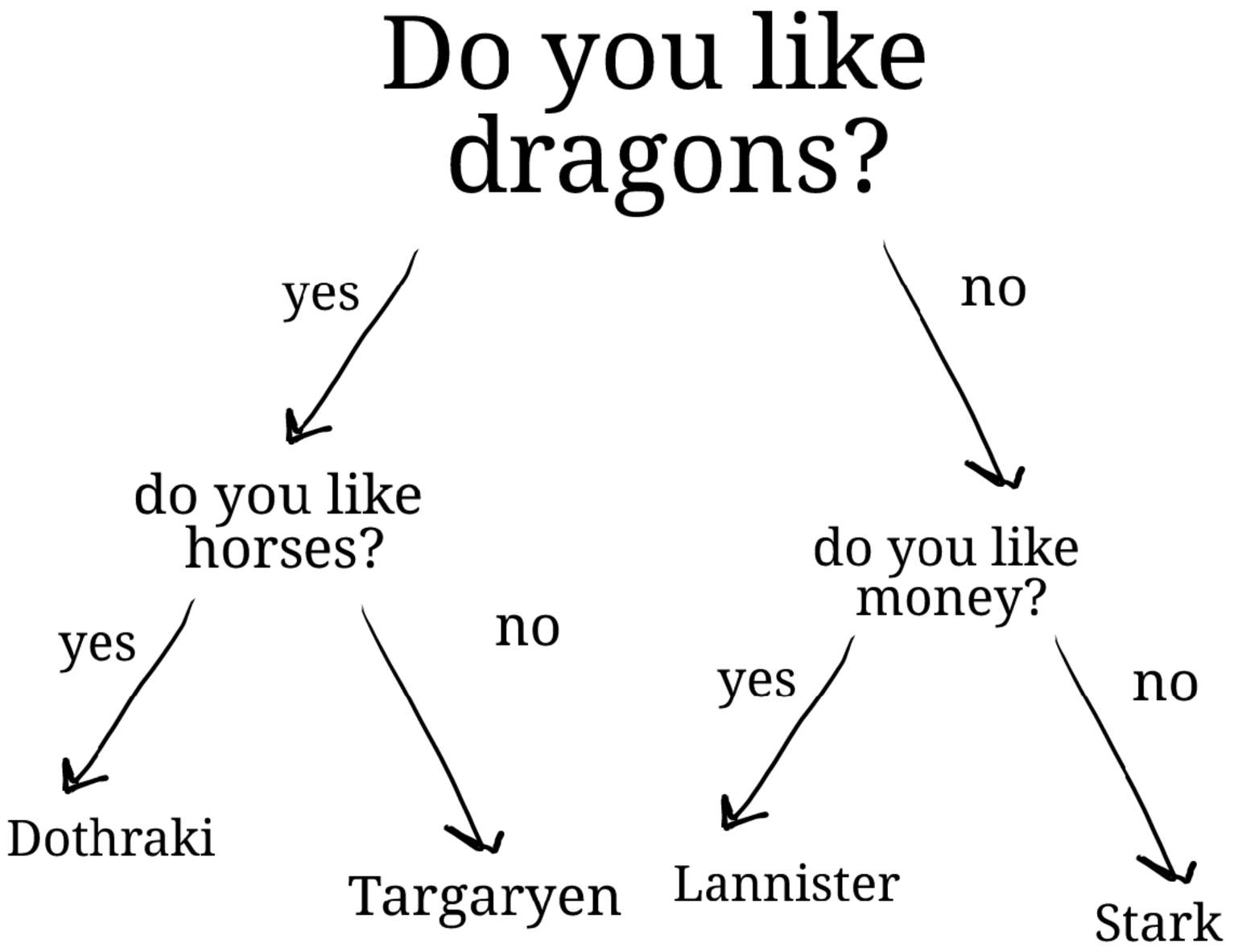

similarity
1
2
3
4
5

Roadmap

- Data
- Clustering
- Classification
- Regression

Features			Class label
Size	Texture	Colour	Fruit
Medium	Crunchy	Red	Apple
Large	Smooth	Green	Apple
Medium	Crunchy	Red	Apple
Medium	Soft	Green	Pear
Large	Soft	Yellow	Pear
Small	Smooth	Yellow	Pear




Do you like dragons?
yes
no
do you like horses?
do you like money?
yes
no
yes
no
Dothraki
Targaryen
Lannister
Stark

Do you like dragons?

yes

no

do you like horses?

do you like money?

yes

no

yes

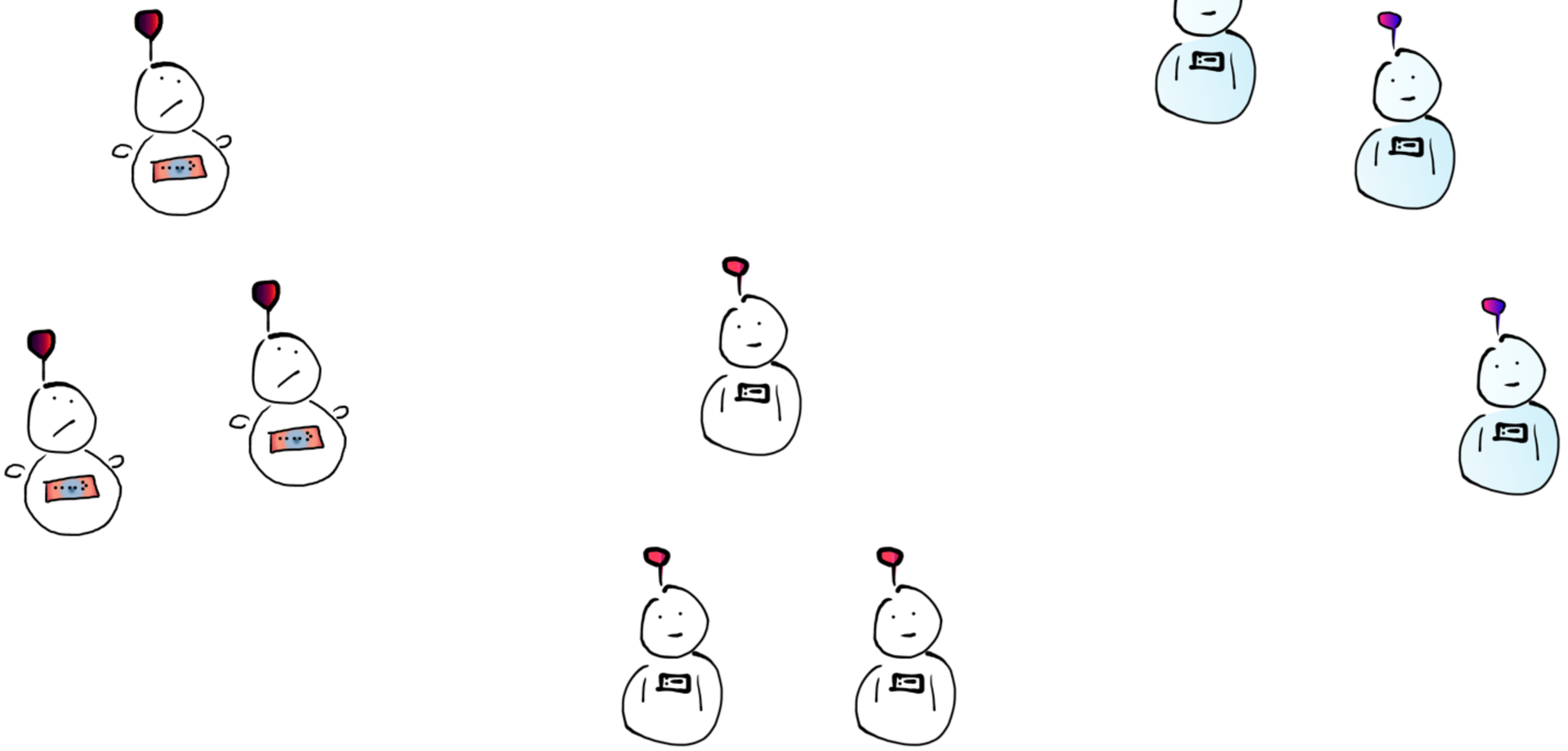
no

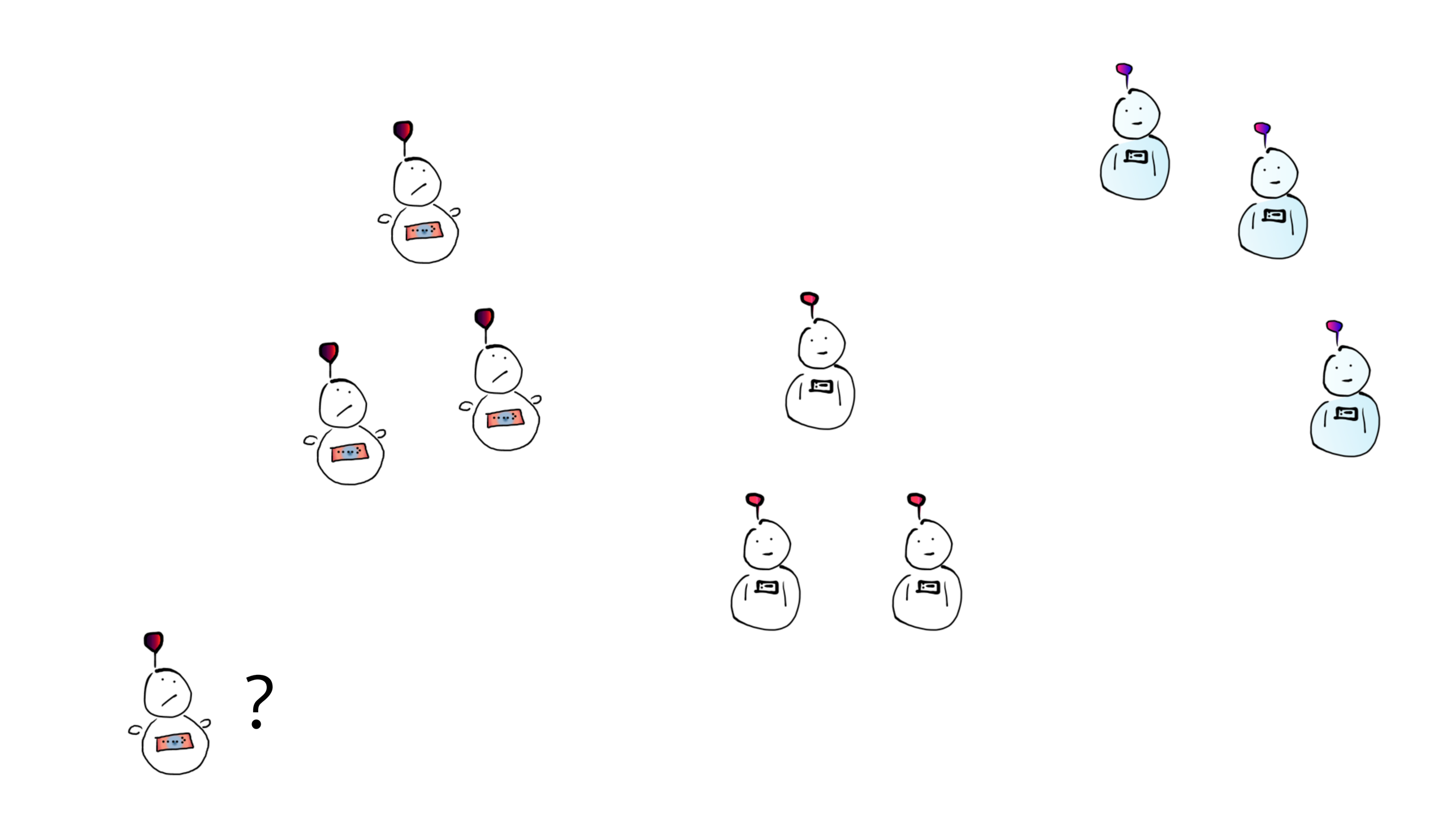
Dothraki

Targaryen

Lannister

Stark





Size	Texture	Colour	Fruit
Medium	Crunchy	Red	Apple
Large	Smooth	Green	Apple
Medium	Crunchy	Red	Apple
Small	Soft	Green	Pear
Large	Smooth	Yellow	Pear

Small	Smooth	Yellow	?

Size	Texture	Colour	Fruit
Medium	Crunchy	Red	Apple
Large	Smooth	Green	Apple
Medium	Crunchy	Red	Apple
Small	Soft	Green	Pear
Large	Smooth	Yellow	Pear

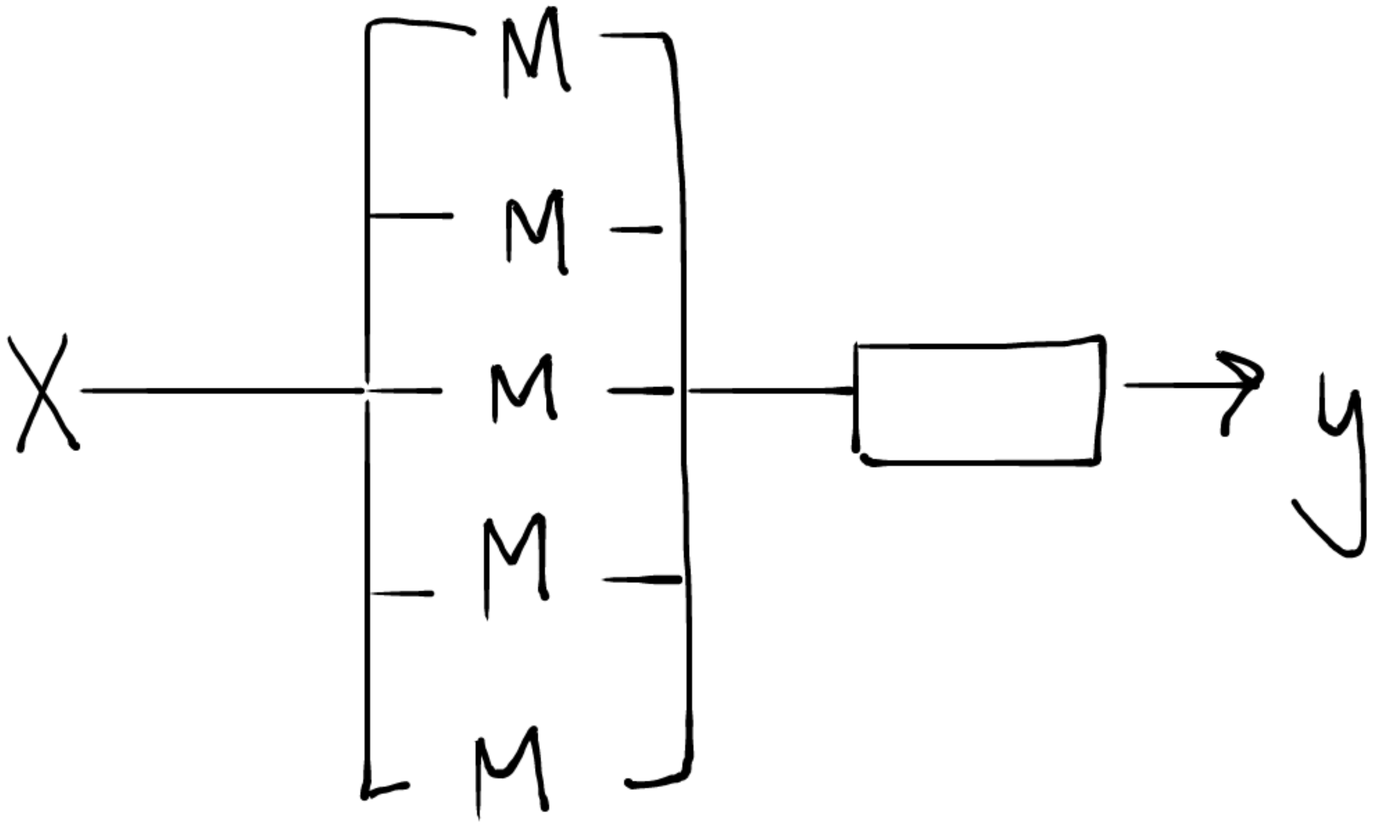
$P(\text{apple}) = \frac{3}{5}$ $P(\text{pear}) = \frac{2}{5}$

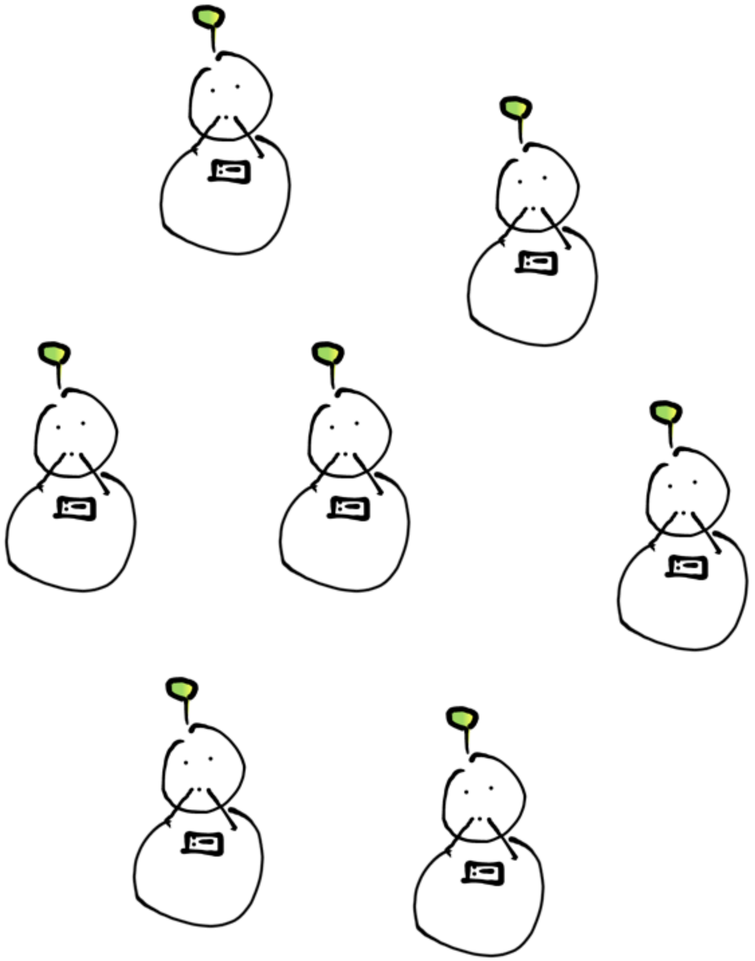
$P(s_s \mid \text{apple}) = 0$ $P(s_s \mid \text{pear}) = \frac{1}{2}$

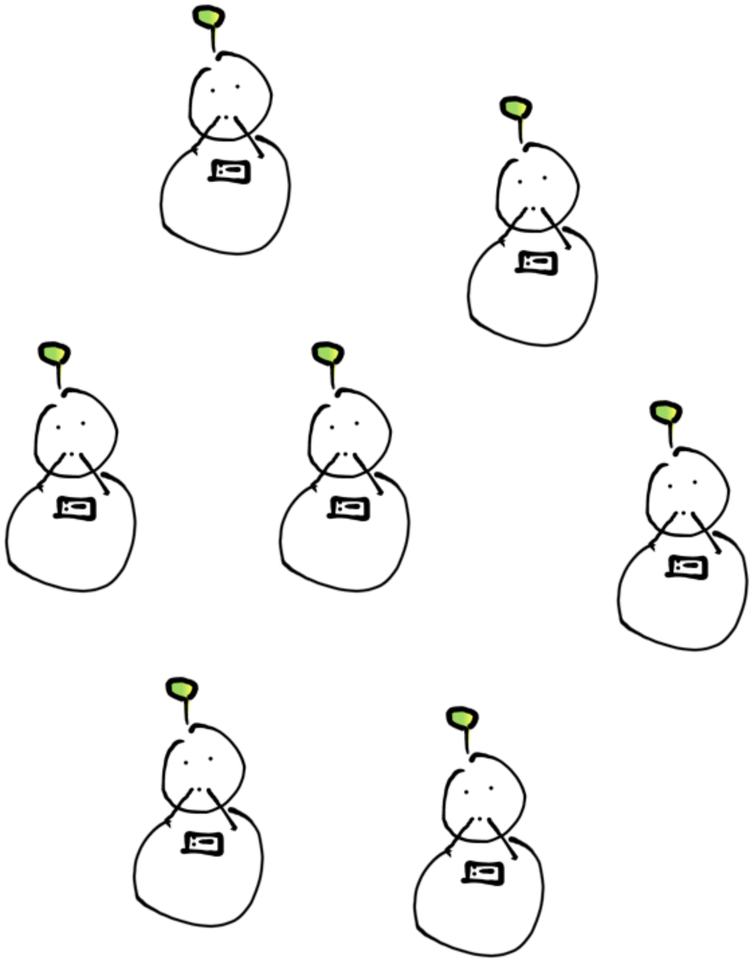
$P(t_s \mid \text{apple}) = \frac{1}{3}$ $P(t_s \mid \text{pear}) = \frac{1}{2}$

$P(c_y \mid \text{apple}) = 0$ $P(c_y \mid \text{pear}) = \frac{1}{2}$

Small	Smooth	Yellow	?


M
M
X
M
M
M
y





Roadmap

- Data
- Clustering
- Classification
- Regression