Fundamentals of Artificial Intelligence

COMP307/AIML420
Machine Learning: fundamentals

Dr. Heitor Murilo Gomes
heitor.gomes@vuw.ac.nz
http://www.heitorgomes.com

Information

- Heitor is teaching until week 4: ML and Search
- If you want to book a time with me, send me an email to arrange a time (*office hours*)
- Extension requests -> submission (ECS) system, other affairs send me an email (heitor.gomes@vuw.ac.nz)
- Helpdesk schedule is on the wiki (starts next week)
- Book Chapters related to the lectures (last lecture: 1.1 and 1.5)
- Students discord server (managed by the class rep) on the wiki

Outline

What is Machine Learning?

Machine Learning tasks

- Supervised (regression and classification)
- Unsupervised (clustering and association rules)

Basic Evaluation

- (Supervised) Train and Test
- Generalisation & overfit

What is machine learning?

"A computer program is said to learn from experience ***E*** with respect to some class of tasks ***T*** and performance measure ***P*** if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***." [1]

[1] Tom Mitchell, *Machine Learning (1997)*

What is machine learning?

"A computer program is said to learn from experience ***E*** with respect to some class of tasks ***T*** and performance measure ***P*** if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***." [1]

Experience ***E*** = Data

[1] Tom Mitchell, *Machine Learning (1997)*

What is machine learning?

"A computer program is said to learn from experience ***E*** with respect to some class of tasks ***T*** and performance measure ***P*** if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***." [1]

Experience ***E*** = Data

Tasks ***T*** = *The ML problem*

[1] Tom Mitchell, *Machine Learning (1997)*

What is machine learning?

"A computer program is said to learn from experience ***E*** with respect to some class of tasks ***T*** and performance measure ***P*** if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***." [1]

Experience ***E*** = Data

Tasks ***T*** = *The ML problem*

Measure ***P*** = *Some metric*

[1] Tom Mitchell, *Machine Learning (1997)*

Machine Learning tasks

Supervised learning*
- Classification and Regression
- Predictive

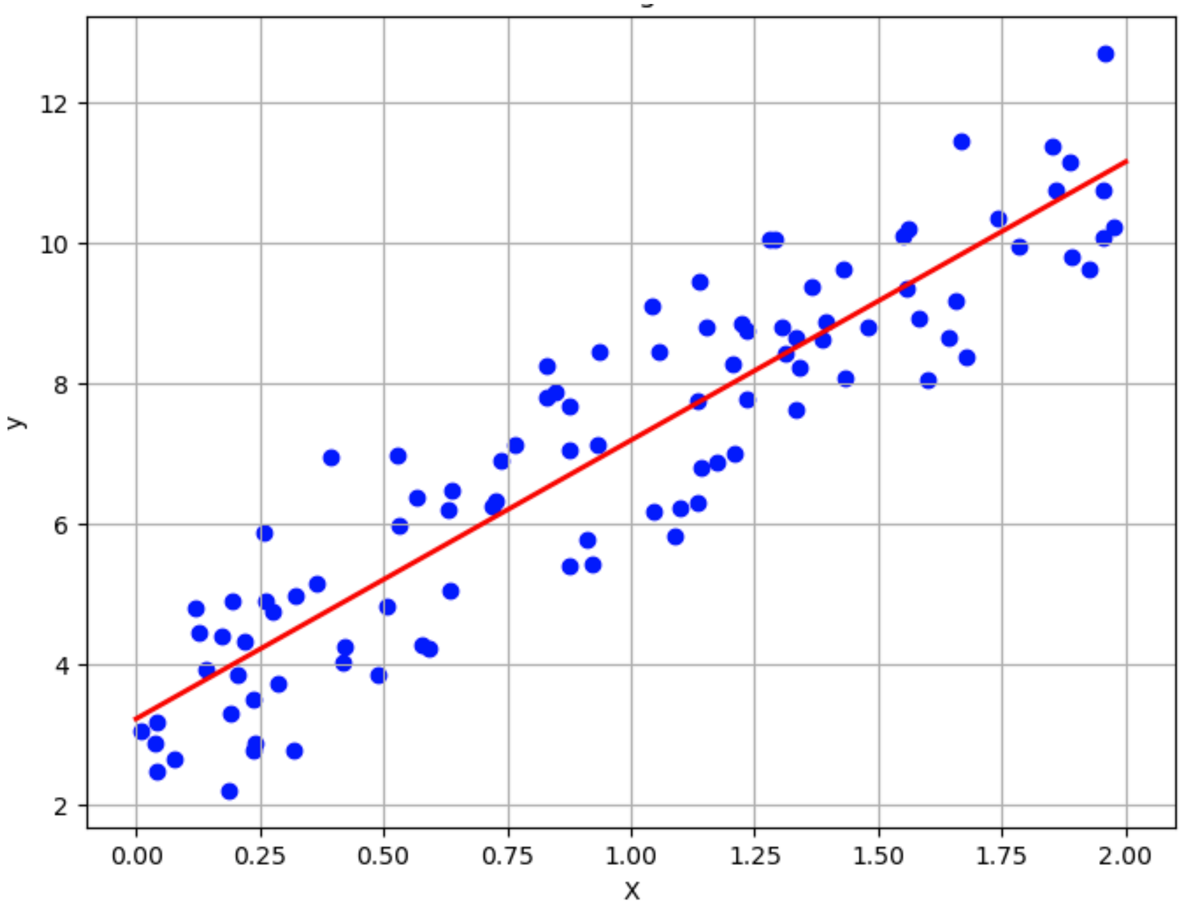
Unsupervised learning*
- Clustering and Association Rules
- Exploratory

Others
- Semi-supervised learning
- Reinforcement learning

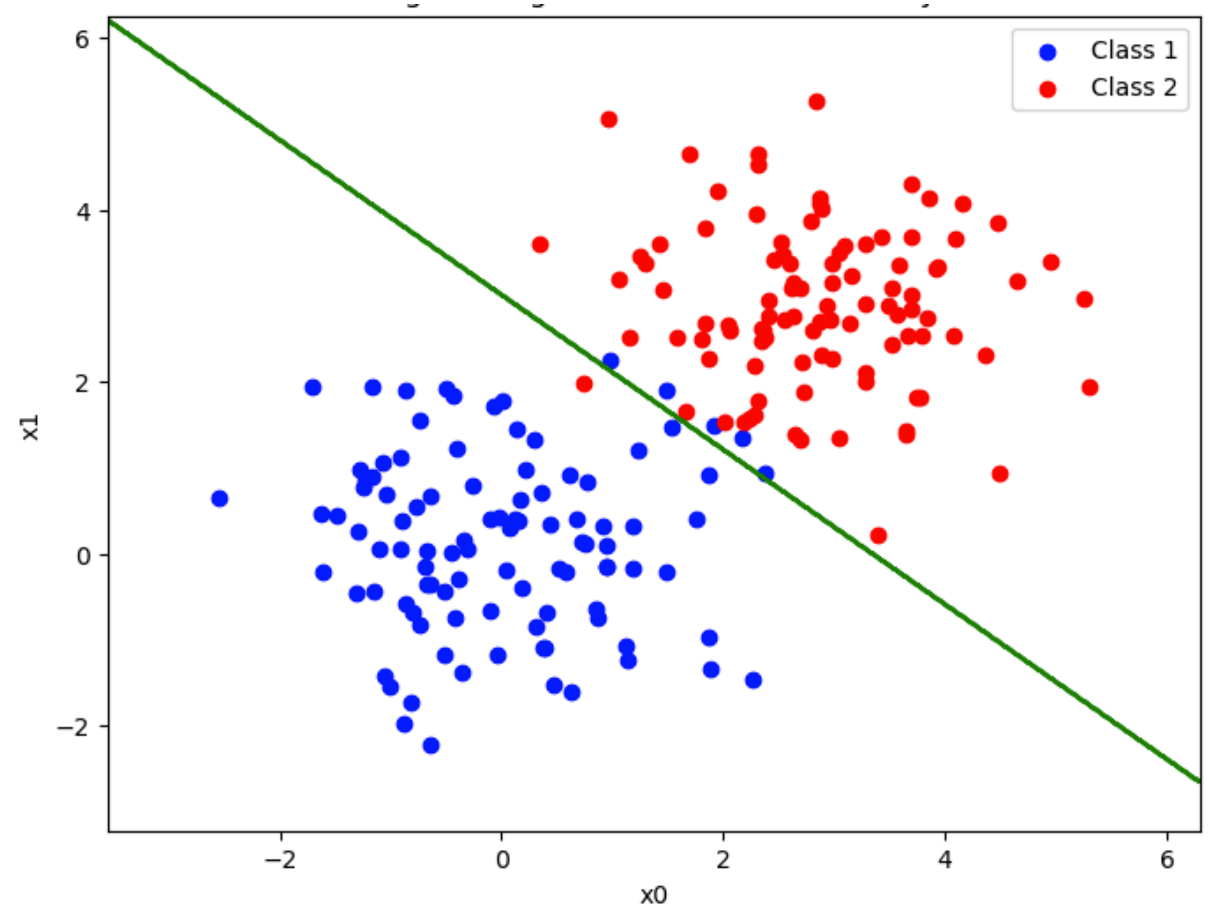
Disclaimer: There are several other tasks that qualify as "supervised" or "unsupervised" for now, let's stick to these for simplicity

Supervised Learning

Regression



Classification



- **Input:** (often) an array of **continuous values**
- **Output:** a **continuous value (target)**

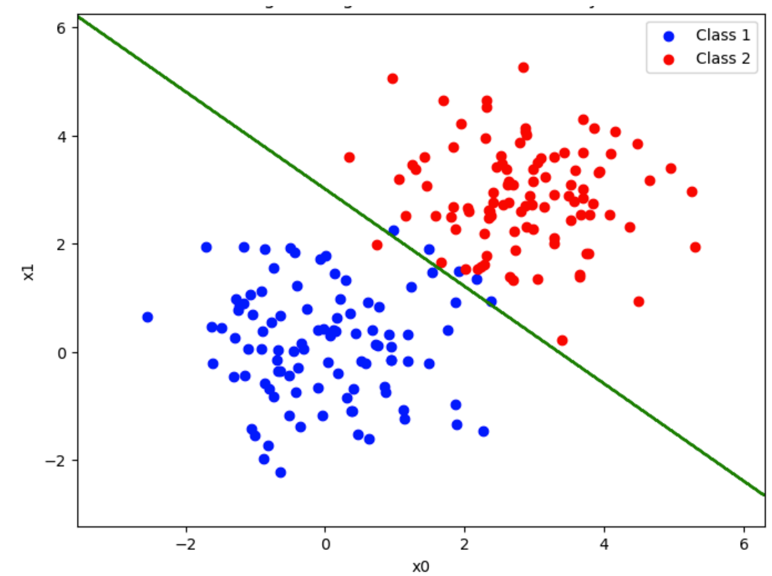
- **Input:** an array of **continuous and/or nominal values**
- **Output:** one of N_c possible **class labels**

Supervised Learning

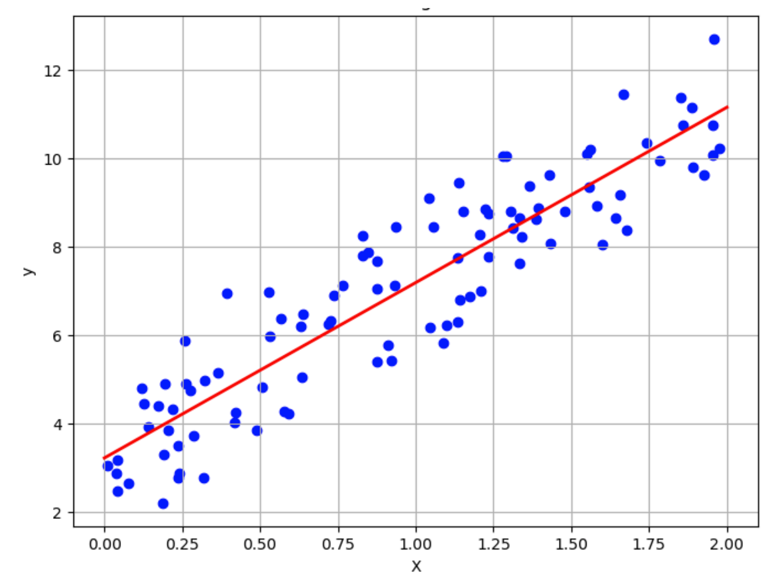
Why ‘supervised’?

There is a **training** (or fitting) phase where **data points** (or instances) with **the ground-truth output** are used to learn the **model.**

Classification



Regression



Supervised Learning

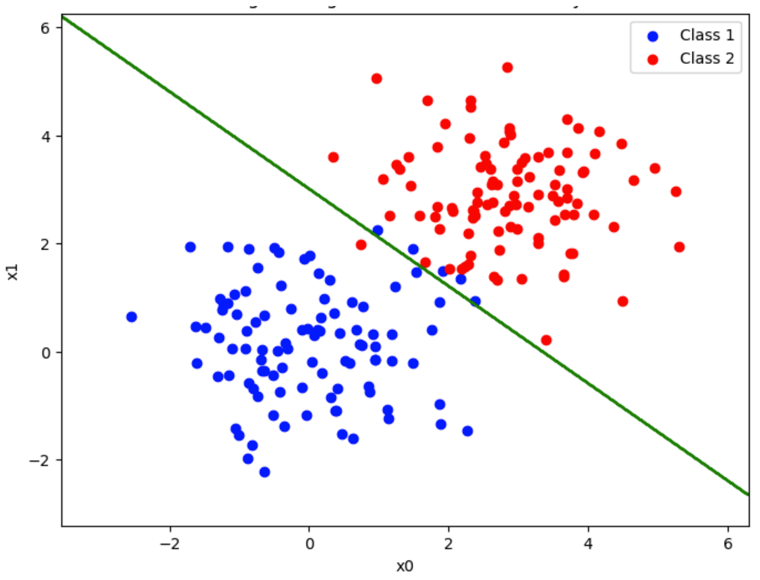
Why 'supervised'?

There is a **training** (or fitting) phase where **data points** (or instances) with **the ground-truth output** are used to learn the **model**.

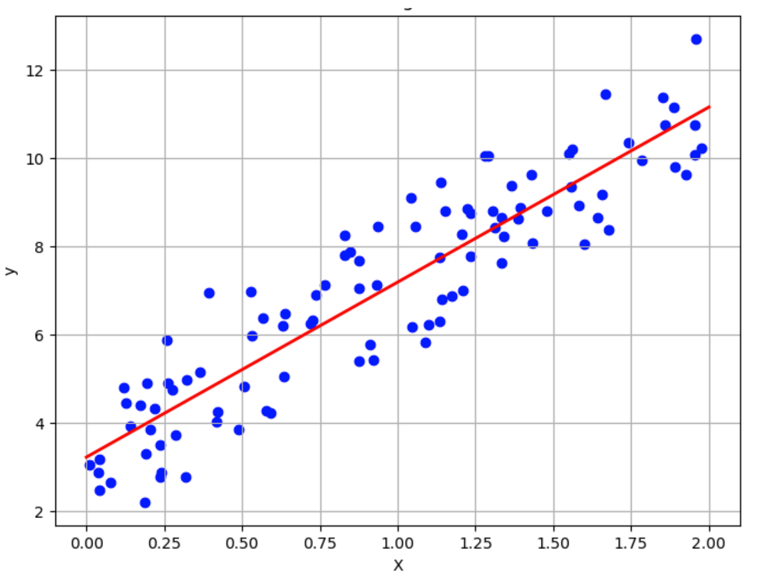
Ground-truth output?

Ground-truth corresponds to the 'correct' output given a particular input. These can be observations "**labeled**" by an expert for example.

Classification



Regression



Supervised Learning

Why ‘supervised’?

There is a **training** (or fitting) phase where **data points** (or instances) with **the ground-truth output** are used to learn the **model**.

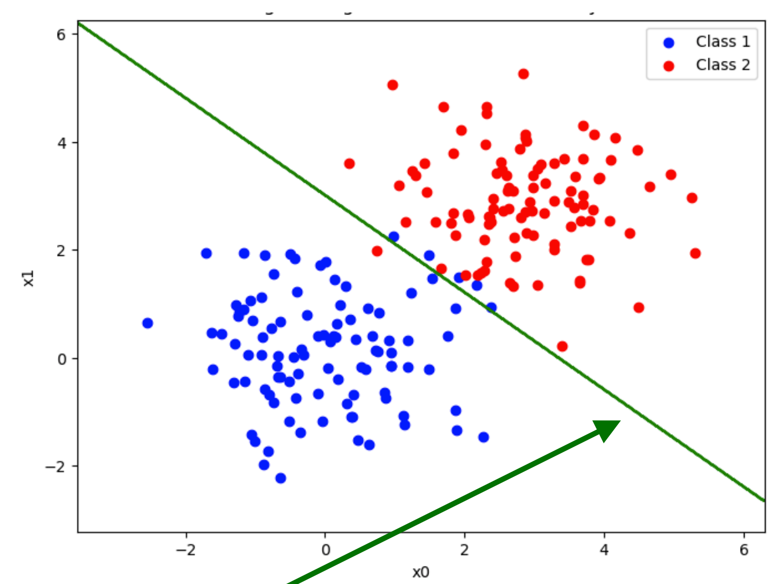
Ground-truth output?

Ground-truth corresponds to the ‘correct’ output given a particular input. These can be observations “**labeled**” by an expert for example.

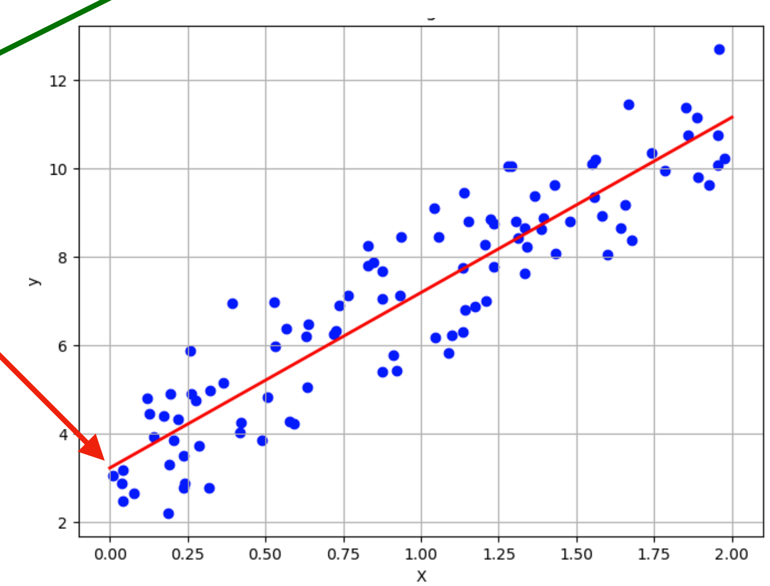
Model?

Model is how we often refer to the structure we obtain once we **train** a particular machine learning algorithm on a data set.

Classification



Regression



Supervised Learning

Why ‘supervised’?

There is a **training** (or fitting) phase where **data points** (or instances) with **the ground-truth output** are used to learn the **model**.

Ground-truth output?

Ground-truth corresponds to the ‘correct’ output given a particular input. These can be observations “**labeled**” by an expert for example.

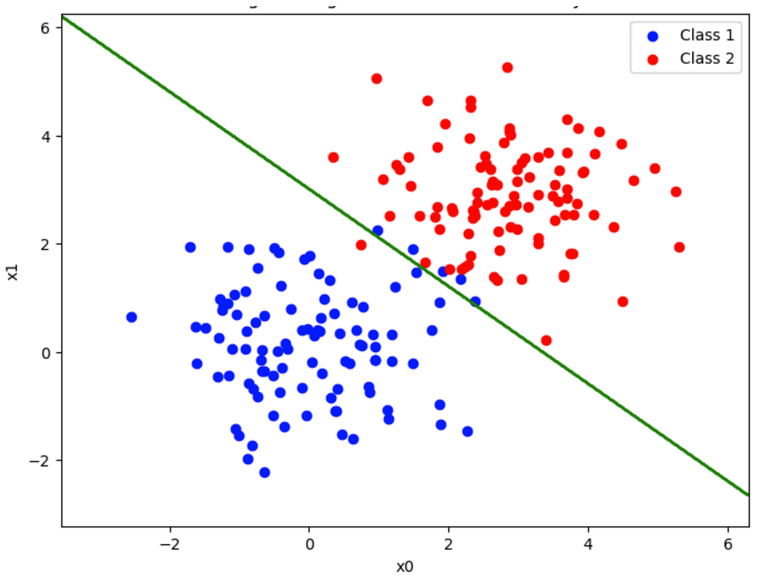
Model?

Model is how we often refer to the structure we obtain once we **train** a particular machine learning algorithm on a data set.

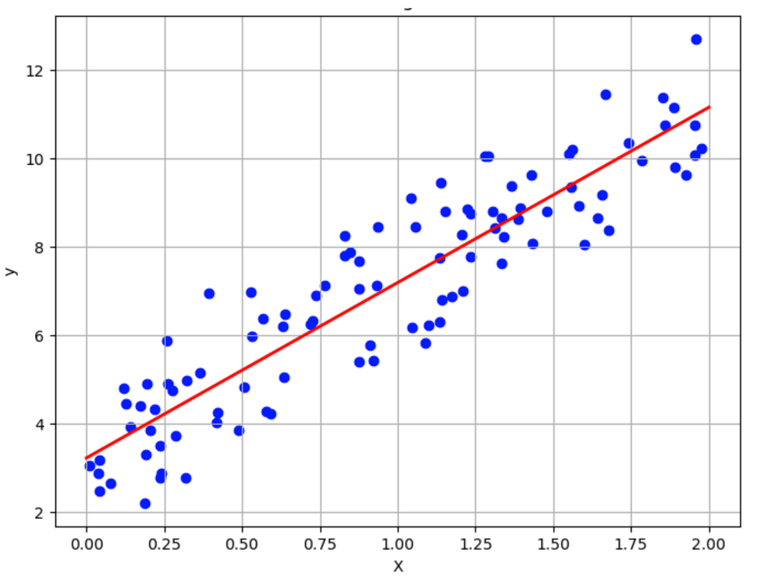
Train?

Training a learner refers to making multiple iterations over a dataset containing the ground-truth to build the **model**.

Classification



Regression



Supervised Learning

Examples

Regression

Predicting the price of a used car based on its age, mileage

Estimating the temperature on a specific day based on historical temperature data and weather factors like humidity and wind speed

Forecasting the sales of a product next month based on historical sales data

Predicting energy consumption based on weather conditions and occupancy

Classification

Identifying whether an email is spam or not spam based on the content of the email and sender information

Classifying handwritten digits (0 to 9) based on the pixel information in an image.

Determining whether a customer is likely to churn (cancel their subscription) based on their past purchase history and demographics

Identifying the genre of a musical piece based on audio features

Supervised Learning

Examples

Regression

Predicting the price of a used car based on its age, mileage

Estimating the temperature on a specific day based on historical temperature data and weather factors like humidity and wind speed

Forecasting the sales of a product next month based on historical sales data

Predicting energy consumption based on weather conditions and occupancy

Classification

Identifying whether an email is spam or not spam based on the content of the email and sender information

Classifying handwritten digits (0 to 9) based on the pixel information in an image.

Determining whether a customer is likely to churn (cancel their subscription) based on their past purchase history and demographics

Identifying the genre of a musical piece based on audio features

$$f(X) \to y$$

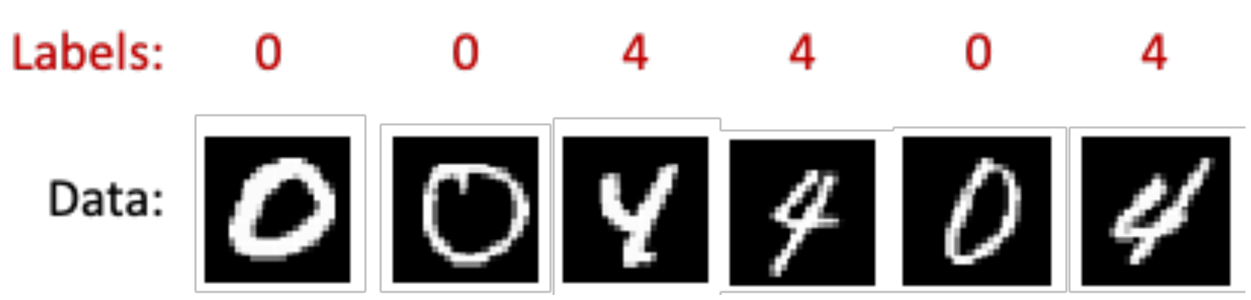
Classification Example

Given some **labeled data**, **<u>train</u>** a machine learning **model** that can **<u>classify</u> new unlabeled data** into **one** of several **class labels**

Classification Example

Given some **labeled data**, **train** a machine learning **model** that can **classify new unlabeled data** into **one** of several **class labels**

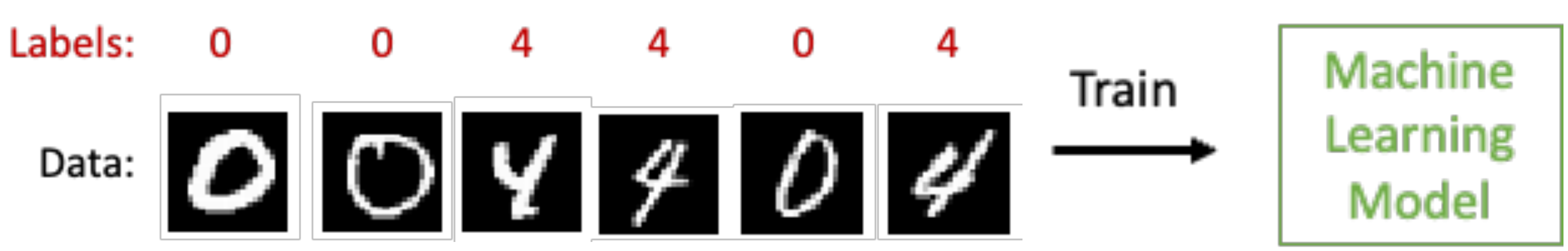
Example: given an image classify whether it is a **0** or a **4**



Classification Example

Given some **labeled data**, **train** a machine learning **model** that can **classify new unlabeled data** into **one** of several **class labels**

Example: given an image classify whether it is a **0** or a **4**



Classification Example

Given some **labeled data**, **train** a machine learning **model** that can **classify new unlabeled data** into **one** of several **class labels**

Example: given an image classify whether it is a **0** or a **4**

Classification
(Definitions)

- Labeled instances (or examples) of the form (x, y)
- $x = x_1, \ldots, x_k$ is a vector of feature (or attribute) values
- y assumes one of N_c possible different class labels
  - Binary classification = 2 classes
  - Multi-class classification > 2 classes

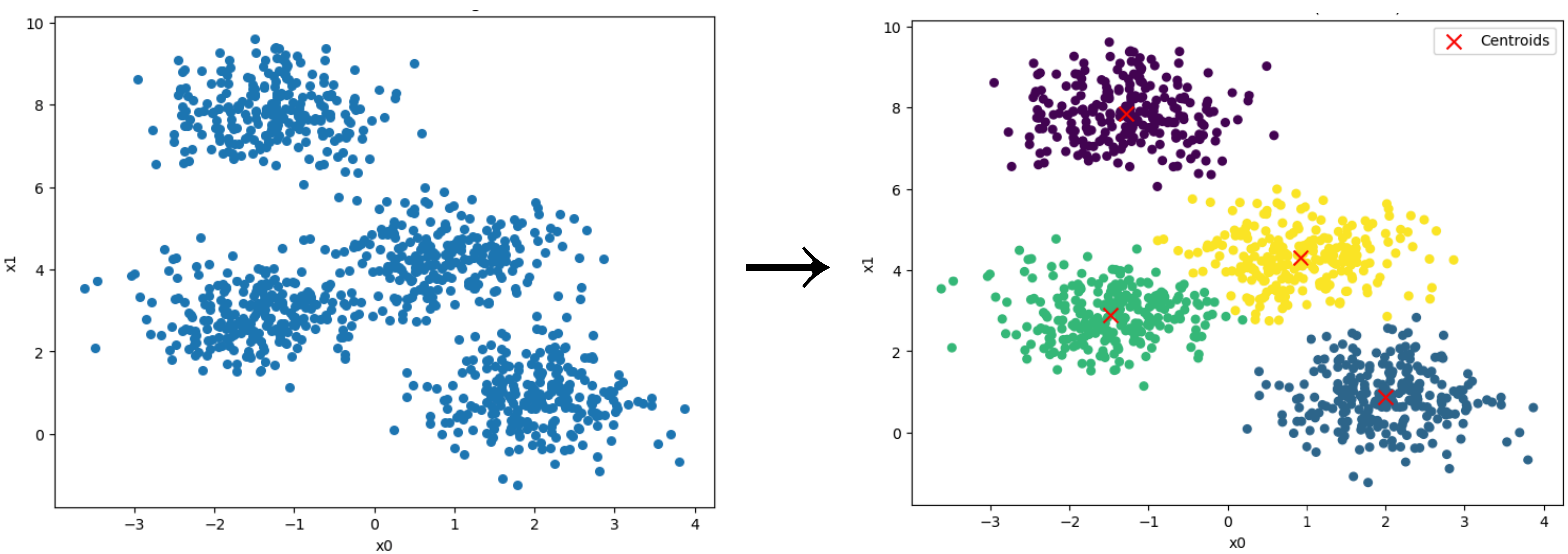
The classifier builds a **model** (or **hypothesis**) f such that $f(x)$ is the predicted class for an unlabeled instance $(x, ?)$

$$f(X) \rightarrow y$$

Unsupervised Learning

- We will focus on **Clustering** and **Association rules**

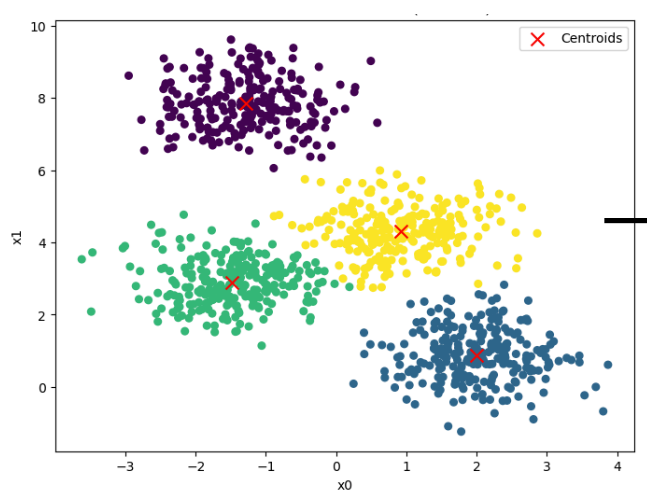
Clustering example



- There are no notion of **class label**, **target** or **ground-truth**

Clustering

- We still have **input data**, but there is **no ground-truth**
- The goal is to discover some **hidden structures**, referred to as **clusters**
- There is no clear “right” answer, requires interpretation



These are not classes!
Meaning is given to them according to an expert

Association Rules

- Classic **data mining*** technique
- They have the form of **if-then statements** that identify **relationships** or dependencies between **items** in a **dataset**
- They reveal **correlations**, not necessarily **causal relationships** between **items**

"**if** *a customer buys bread*, they are **also likely** to buy *milk*"

or simply:

$$bread \rightarrow milk$$

Association Rules

- Classic **data mining*** technique

- They have the form of **if-then statements** that identify **relationships** or dependencies between **items** in a **dataset**

- They reveal **correlations**, not necessarily **causal relationships** between **items**

"**if** *a customer buys bread*, they are **also likely** to buy *milk*”

or simply:

$$bread \rightarrow \boxed{milk}$$

Antecendent **Consequent**

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	marmite	butter	banana
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

Market basket analysis

Each **transaction** may represent something that one customer bought at once

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	marmite	butter	banana
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

Market basket analysis

Each **transaction** may represent something that one customer bought at once

Some possible rules

$$bread \rightarrow marmite$$

$$apple, bread \rightarrow banana$$

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	marmite	butter	banana
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

Market basket analysis

Each **transaction** may represent something that one customer bought at once

Some possible rules

$$bread \rightarrow marmite$$

$$apple, bread \rightarrow banana$$

How do we "quantify" these rules?

Association Rules

- To quantify (and probably rank) the rules, we commonly use their **Support** and **Confidence**

- **Support:** the **proportion of transactions** where both the antecedent and consequent appear together

- **Confidence:** the **likelihood** of finding the consequent in a transaction **given** the presence of the antecedent

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	banana	butter	marmite
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

$$bread \rightarrow marmite$$

Support = # bread and marmite together / total transactions

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	banana	butter	marmite
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

$$bread \rightarrow marmite$$

Support = # bread and marmite together / total transactions

Support = 5 / 10 = 0.5

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	banana	butter	marmite
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

$$bread \rightarrow marmite$$

Support = # bread and marmite together / total transactions

Support = 5 / 10 = 0.5

Confidence = # marmite and bread / # bread in transactions

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	banana	butter	marmite
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

$$bread \rightarrow marmite$$

Support = # bread and marmite together / total transactions

Support = 5 / 10 = 0.5

Confidence = # marmite and bread / # bread in transactions

Confidence = 5 / 8 = 0.625

Association Rules

Example

Transaction 1	bread	marmite	milk	
Transaction 2	bread	butter	banana	
Transaction 3	milk	apple		
Transaction 4	bread	banana	butter	marmite
Transaction 5	bread	marmite	apple	
Transaction 6	bread	marmite	vegemite	apple
Transaction 7	bread	marmite	butter	vegemite
Transaction 8	bread	milk	vegemite	
Transaction 9	milk	banana		
Transaction 10	bread	milk	butter	banana

$$bread \rightarrow marmite$$ (**Support** = 0.5 , **Confidence** = 0.625)

$$apple, bread \rightarrow banana$$ (**Support** = 0.0 , **Confidence** = 0.0)

$$bread \rightarrow vegemite$$ (**Support** = 0.3 , **Confidence** = 0.375)

Association Rules
Conclusions

- There is a relationship between confidence of $A \rightarrow B$ and the conditional probability $P(B \mid A)$
- There are several algorithms to mine association rules, such as Apriori [2]
- Specify thresholds for the support and confidence to reduce number of rules
- Applications:
  - Optimize product placement, recommendation, diagnosis, …

Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215. 1994.

Other tasks?

- **Semi-supervised learning**
  - Use unsupervised to “help” on a classification problem, and vice-versa
- **Reinforcement learning**
  - Given a sequence of **actions** and **states**, and **reward/ penalty**
  - Infer a **policy** for choosing best **actions**

Basic evaluation
Supervised learning

Why evaluate?

Basic evaluation

Supervised learning

"A computer program is said to learn from experience ***E*** with respect to some class of tasks ***T*** and performance measure ***P*** if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***." [1]

Experience ***E*** = Data
Tasks ***T*** = *The ML problem*
Measure ***P*** = *Some metric*

[1] Tom Mitchell, *Machine Learning (1997)*

Why evaluate?

To answer the questions:

(1) Has this model learned anything?

(2) Between model A and model B, which one is best?

Basic evaluation
Supervised learning

How do we evaluate?

There are two dimensions to that:

- the **procedure;**
- and the **metric(s)**

Basic procedure: split the data into **training** and **testing**

Basic evaluation

Supervised learning

Why **training** and **testing**?

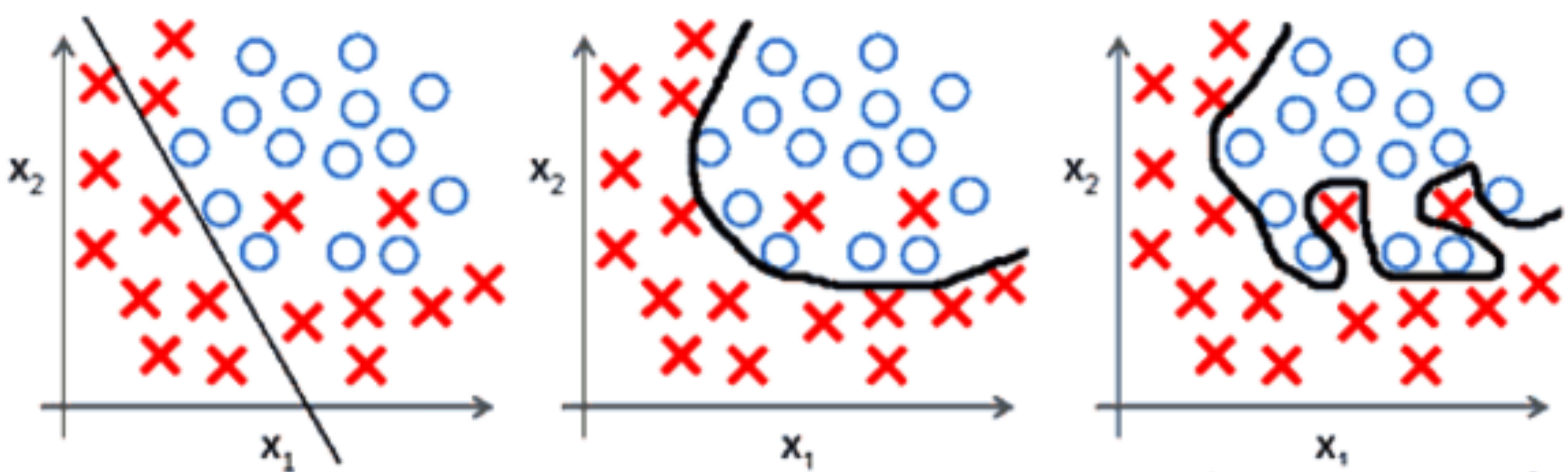
- Learn the model on one portion (**train**) of the data and verify if it **“generalises”** to the other portion (**test**)

Basic evaluation

Supervised learning

Why **training** and **testing**?

- Learn the model on one portion (**train**) of the data and verify if it **“generalises”** to the other portion (**test**)



Assuming these are the decision boundaries of three models on the **train** data. **Which model you think is “better”**?

Wrap-up

- **Discussed the high-level characteristics of supervised and unsupervised learning**

- Talked a bit more about **Association Rules** as we won't be seeing it again

- **What about evaluating unsupervised algorithms?**

    - It is much more complicated, but we will discuss this for clustering algorithms

- **What about algorithms for Classification and Clustering?**

    - We have specific lectures for them, do not worry!

Wrap-up cont.

- Chapters associated with this lecture [3]: 19.1 and 19.2 (may want to check 19.3.4)

Next lecture:

- kNN (classification algorithm) and more on evaluation (including k-fold CV)

[3] Russell, Stuart J., and Peter Norvig. *Artificial intelligence a modern approach*. 4th edition