Introduction to Artificial Intelligence

1897

VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

COMP307/AIML420

Machine Learning 1: Fundamentals

Dr Andrew Lensen

Andrew.Lensen@vuw.ac.nz

Announcements

- Alert level 2 this week
  - Scan the QR code on your way out if you haven’t!
- Don’t panic, but stay home if feel unwell and ring healthline: (0800) 358 5453
- If you have a mask – please wear it!
  - Not sitting at a sticker? You must wear one
- All helpdesks (+ office hours) are online this week
  - https://vuw.zoom.us/my/comp307

Outline

- Why Machine Learning?
- What is machine learning?
- Types of machine learning
- Machine learning algorithms
- Training set vs test set
- Generalisation

Why Machine Learning (ML)?

- To make smarter machines (systems)
  - Improve performance, without (or with little) human intervention
  - Robust behavior in noisy environments
  - “Learn about the world” in order to act sensibly

- Digit recognition, face recognition, …
- Automatic software testing, anomaly detection
- Robot soccer, AlphaGo, …
- Automatic paper writing, music composing, …

- To understand intelligence

- Because it’s interesting!

What is Machine Learning (ML)?

- Machine learning is concerned with the design and development of algorithms and techniques that allow computers to “learn”
- “Machine learning is the study of computer algorithms that improve automatically through experience”
- Any system which changes itself
- Any system which improves its performance over time
- “Making sense of the world”
- “Finding patterns and commonalities in experience”
- …

Two Approaches

- Using ML to build/train intelligent agents (offline learning)
  - Building an expert system by training on pre-classified examples
  - Building a voice recognition system by training on large datasets
  - Building a face detection system by training on a face dataset
  - *Agent does* ***NOT*** *learn while working, learning can be very slow*

- Building agents that learn from experience and improve their performance over time (online learning)
  - Spam filtering system that learns from ongoing user feedback
  - Household robot that learns what the owners want
  - *Agent learns while working, learning* ***must*** *be fast*

Inputs and Outputs of Learning Systems

- What is being learned (and how is it represented)?
  - Classifiers / Predictors
  - Concept descriptions
  - Models of the world
  - Rules for choosing actions
  - (Hidden) patterns / features

- What is it learned from? (and how is it represented)?
  - Set of instances
  - Sequence of actions / states
  - Labeled / unlabeled / reward
  - Batch or incremental

Types of Learning Systems

One helpful categorisation:

- **Supervised** learning
- **Unsupervised** learning
- (**Semi-supervised** learning)
- **Reinforcement** learning

Supervised Learning

- **Given**: instances of inputs and target outputs (**labels**)
- **Generate**: a function that maps inputs to desired outputs
- **Predict**: the correct output for a **new (unseen)** input
- Examples:
  - Learn rules for mortgage approval from records of past decisions
  - Learn to recognise words from handwriting documents
  - Learn a model or rule for postal(zip) code recognition
  - Learn patterns/trends for predicting the stock market/weather/traffic
  - Learn patterns/features from fingerprints to detect terrorists at airports

- Most widely explored type of machine learning (*for now…*)
- Many different approaches

Other Learning Types

- **Unsupervised Learning**
  - Given: set of **unlabelled** instances
  - Generate: knowledge around the underlying structure of the data
  - Examples:
    - Find clusters in high-dimensional data
    - Construct species hierarchy
    - Group search engine results into categories to refine a search query
    - Identify parts of genes that have similar properties
- **Semi-supervised learning**:
  - A *mixture* of supervised learning and unsupervised learning
- **Reinforcement Learning**:
  - Given: sequence of actions and states, occasional reward/penalty
  - Generate: policy for choosing best actions
  - Examples: Robot navigation tasks, Multiple lift controller, ...

Machine Learning Tasks

Supervised

- **Classification/Prediction**
- **Regression**
- …

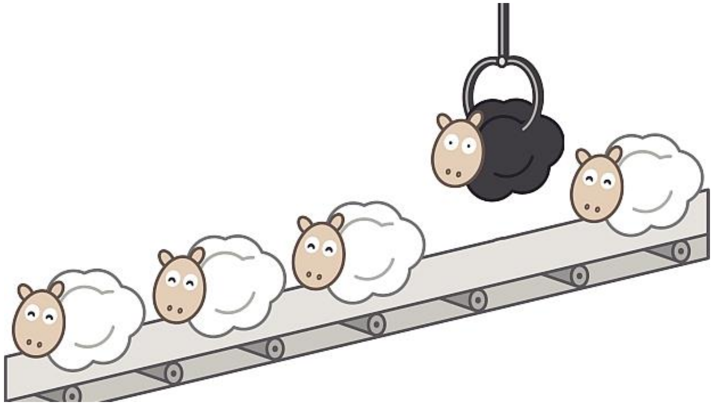
Unsupervised

- **Clustering**
- **Association Rule Mining** (Link analysis)
- …

Classification

- Maps data into predefined groups (classes)
- *Supervised* learning
- Need labelled data in advance

- Examples
  - Medical: cancer vs not cancer
  - Bank: credit reliable vs unreliable
  - Digit recognition: *multi-class*
  - Weather: sunny or rainy (Boolean)
  - Anomaly detection
  - …



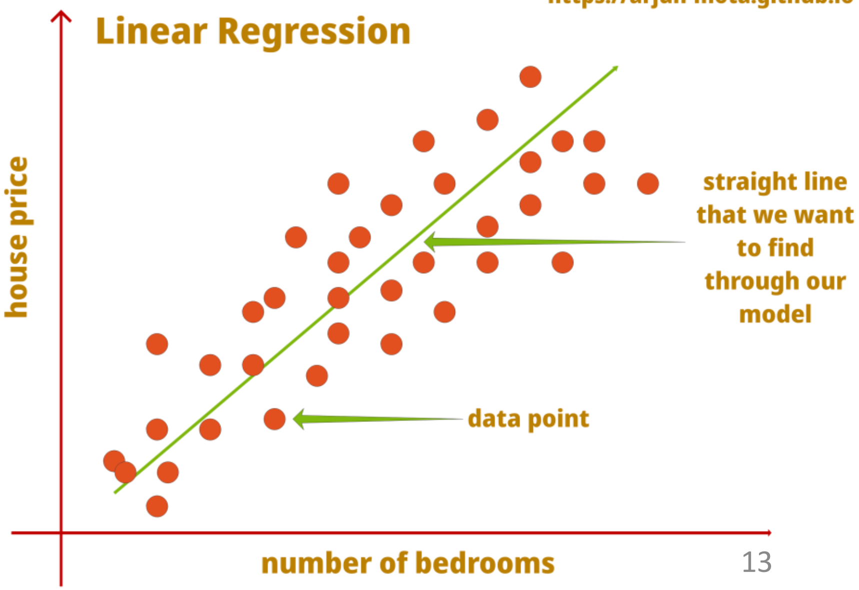
Regression

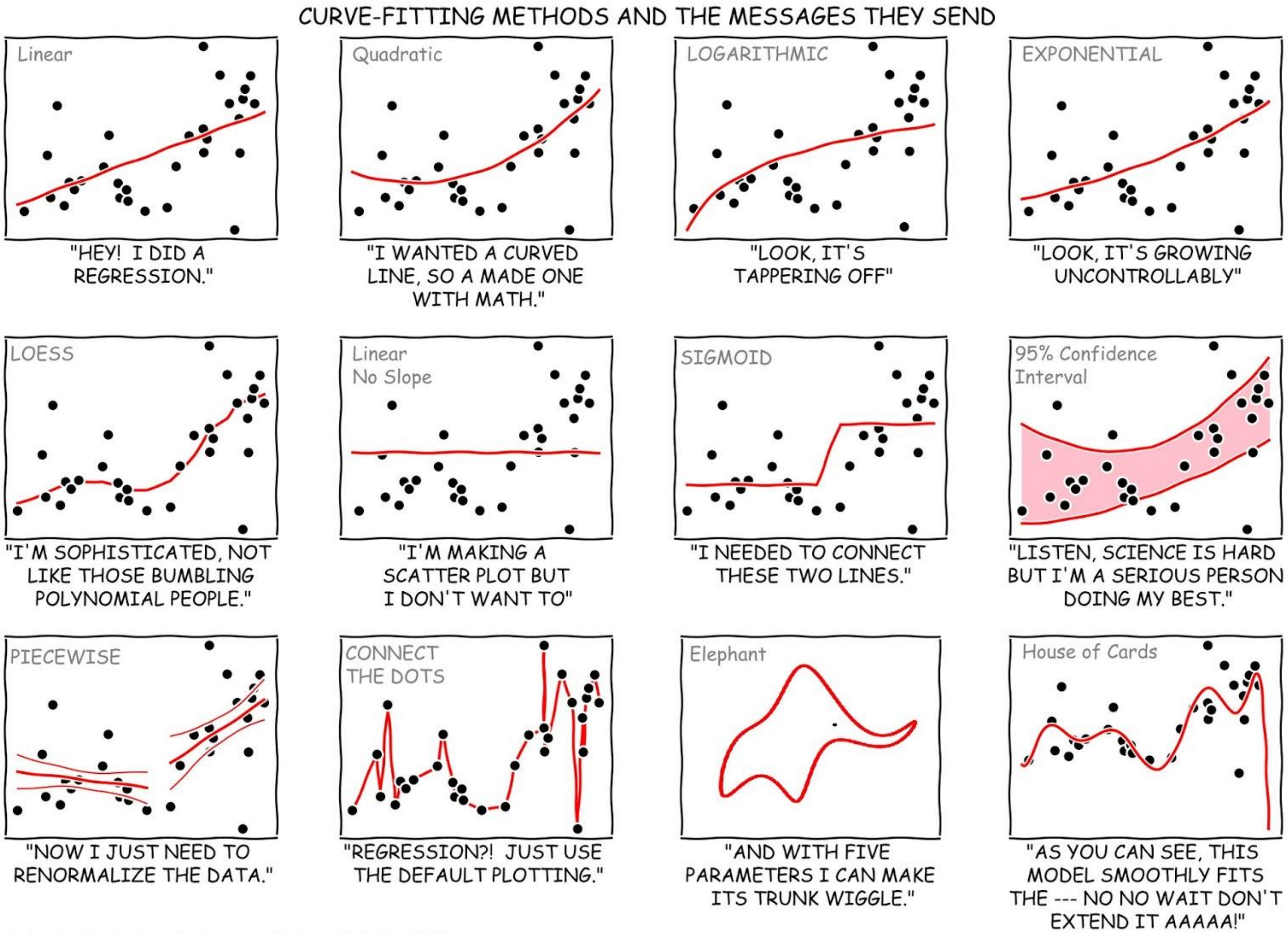
将数据项映射到实值预测变量

- Map a data item to a real-valued prediction variable
- Supervised learning
- Learning a function
- Often assume a certain function type (e.g. linear, logistic, polynomial, …) and determine the best function of this type to fit the given data
- Or, learn the function type at the same time (Symbolic Regression)

- Examples
  - Financial prediction
  - Saving prediction
  - Ad cost vs sales

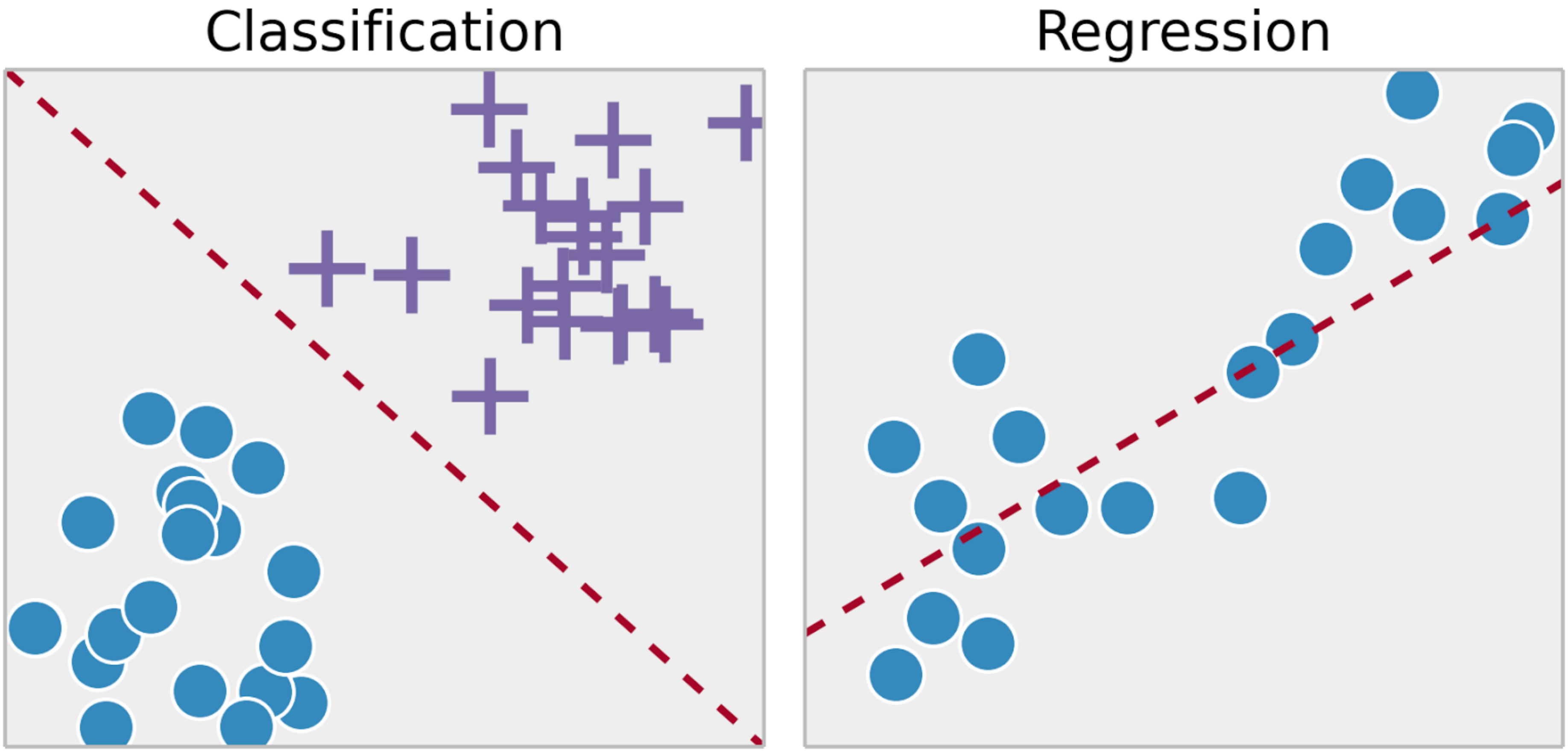
https://arjun-mota.github.io




CURVE-FITTING METHODS AND THE MESSAGES THEY SEND
Linear
"HEY! I DID A REGRESSION."
Quadratic
"I WANTED A CURVED LINE, SO A MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPPERING OFF"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
Linear No Slope
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO"
SIGMOID
"I NEEDED TO CONNECT THESE TWO LINES."
95% Confidence Interval
"LISTEN, SCIENCE IS HARD BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"NOW I JUST NEED TO RENORMALIZE THE DATA."
CONNECT THE DOTS
"REGRESSION?! JUST USE THE DEFAULT PLOTTING."
Elephant
"AND WITH FIVE PARAMETERS I CAN MAKE ITS TRUNK WIGGLE."
House of Cards
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE --- NO NO WAIT DON'T EXTEND IT AAAAA!"

by Douglas Higinbotham in Python inspired by https://xkcd.com/2048

Classification vs Regression



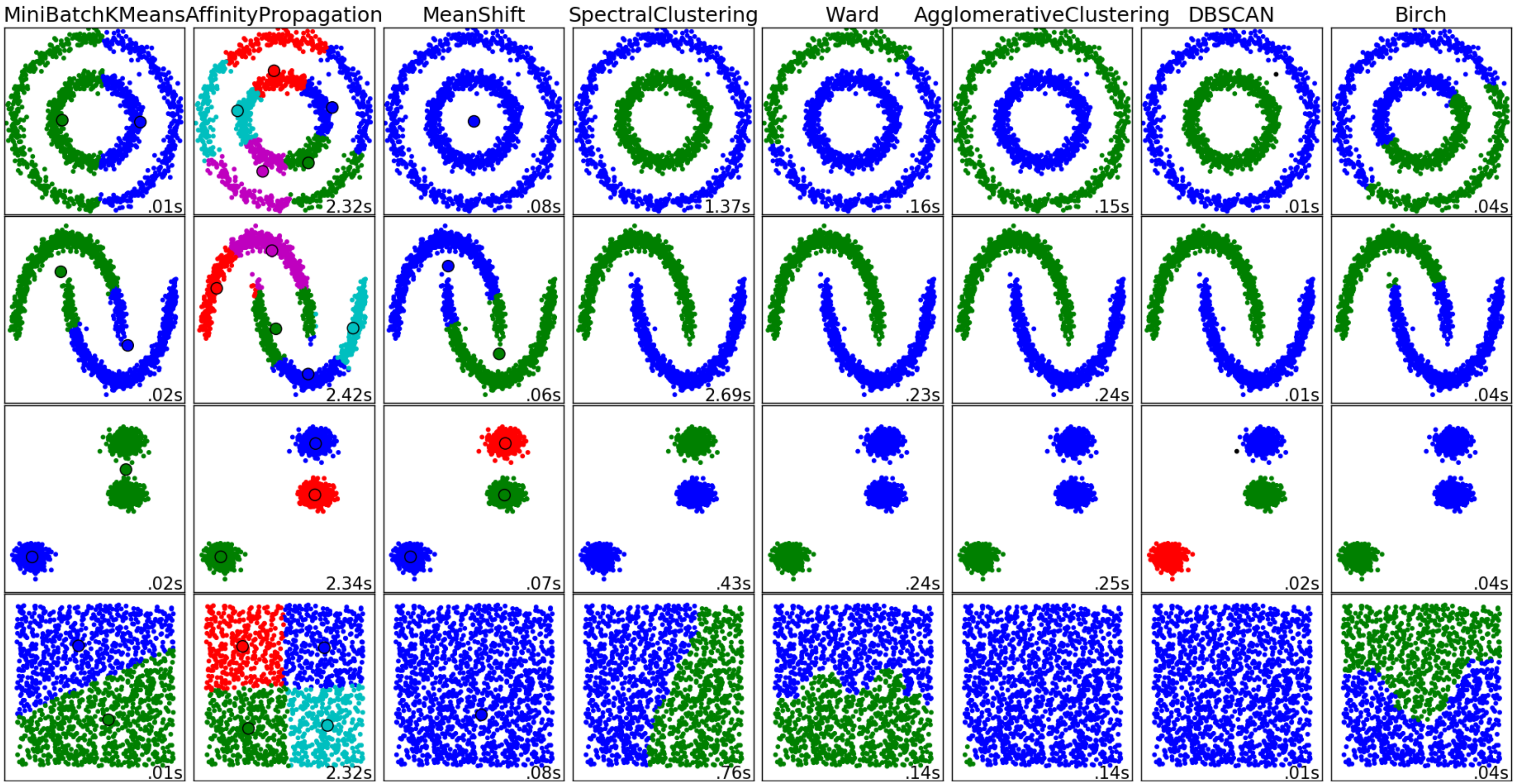
Clustering

- Unlike classification, the groups are not predefined, but rather defined by the data itself (no class labels!)
- Unsupervised learning
- Segmenting or partitioning data into groups that might or might not be disjointed
- Done by determining the similarity/distance among the data on predefined attributes
- A domain expert is needed to interpret the meaning

将数据分段或分区到可能脱节也可能不脱节的组中
通过确定预定义属性上的数据之间的相似性/距离来完成
需要领域专家来解释其含义

集群:主观!

Clustering: *subjective*!



Association Rules

- Link analysis = association

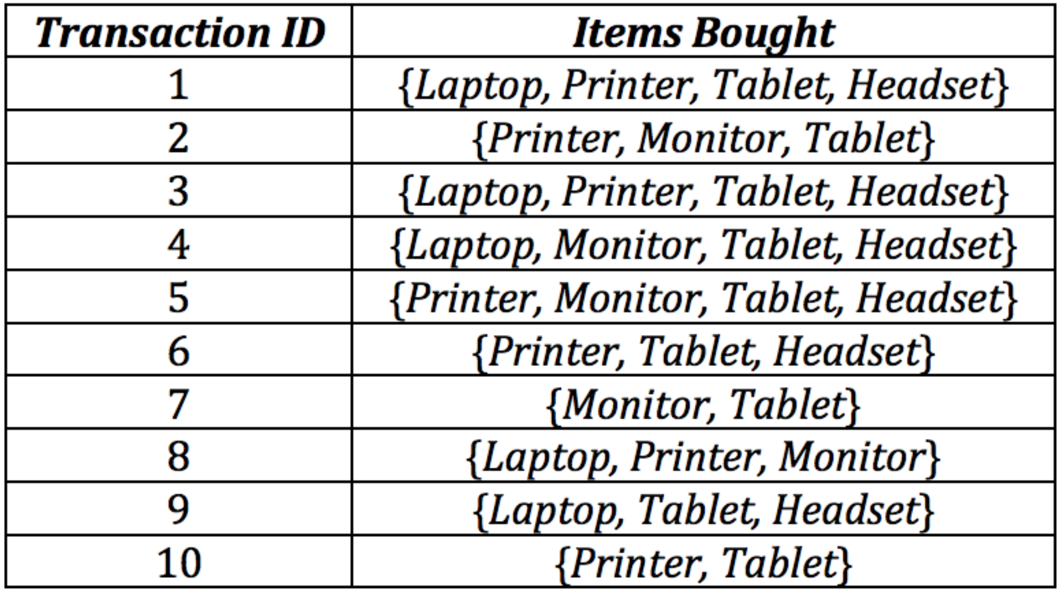
链接分析 == 关联
发现数据之间的关系

- Uncover relationships among data
- An association rule is a model that identifies specific types of data associations

关联规则是一种模型，用于识别特定类型的数据关联

- Often used in the retail sales community to identify items that are frequently purchased together

通常在零售社区中用于识别经常一起购买的物品



Transaction ID	Items Bought
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

Association Rules: Beer & Nappies!

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
…	…

market basket transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

- Probably just a nice anecdote!
- http://www.dssresources.com/newsletters/66.php

Main Learning Paradigms/Techniques

- **Case-based learning** (or instance-based learning): Use specific cases or experiences and rely on flexible matching methods to retrieve similar cases.
  - Example: K-*nearest neighbour (next lecture!)*
- **Induction learning**: Induce a general rule from a set of examples
  - Example: *decision trees (next week!)*
- **Statistical (probability based) learning**:
  - *Naive Bayes (second half!)*
  - *Support Vector Machines*
  - *Bayesian Belief Networks* (AIML429)
- **Analytic learning systems**: Represent knowledge as rules in logic form
  - Example: *Horn clauses*

Main Learning Paradigms/Techniques

- **Connectionist learning**: based on human brain behaviour
  - – *artificial neural networks (AIML425)*

- **Genetic/evolutionary learning**: based on the mechanism of natural selection and natural genetics. (AIML426)
  - – *Genetic algorithms*: evolve *bit strings* or *chromosomes*
  - – *Genetic programming*: evolve computer programs
  - – *PSO, EMO, LCS, …*

- **Hybrid learning…**

Supervised Learning Systems

- **Simple** systems:
  - Representation: feature vectors
  - no missing values
  - no errors
  - sufficient features and sufficient examples

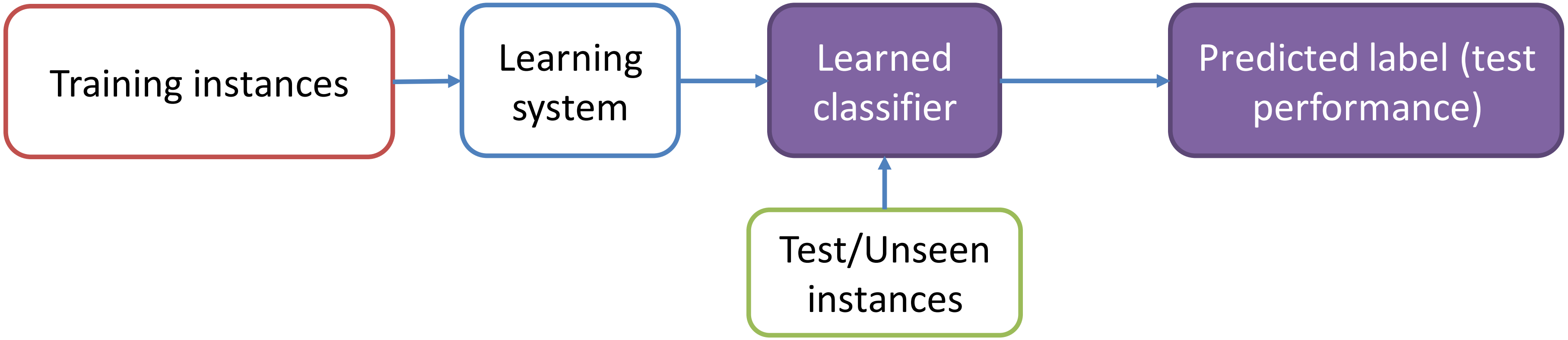
Length	Width	Height	Colour	Class
96.5cm	40.6cm	15.2cm	Brown	Guitar

- **Complex:**
  - Representation: multiple components and relationships
  - missing values
  - noisy data
  - limited examples

Length	Width	Height	Colour	Class
965cm	?	15.2cm	True	Guitar

A Typical Supervised Learning System

- Presented with a set of training instances, some positive and some negative
- Need to come up with a rule/pattern that distinguishes the positive examples from the negative ones

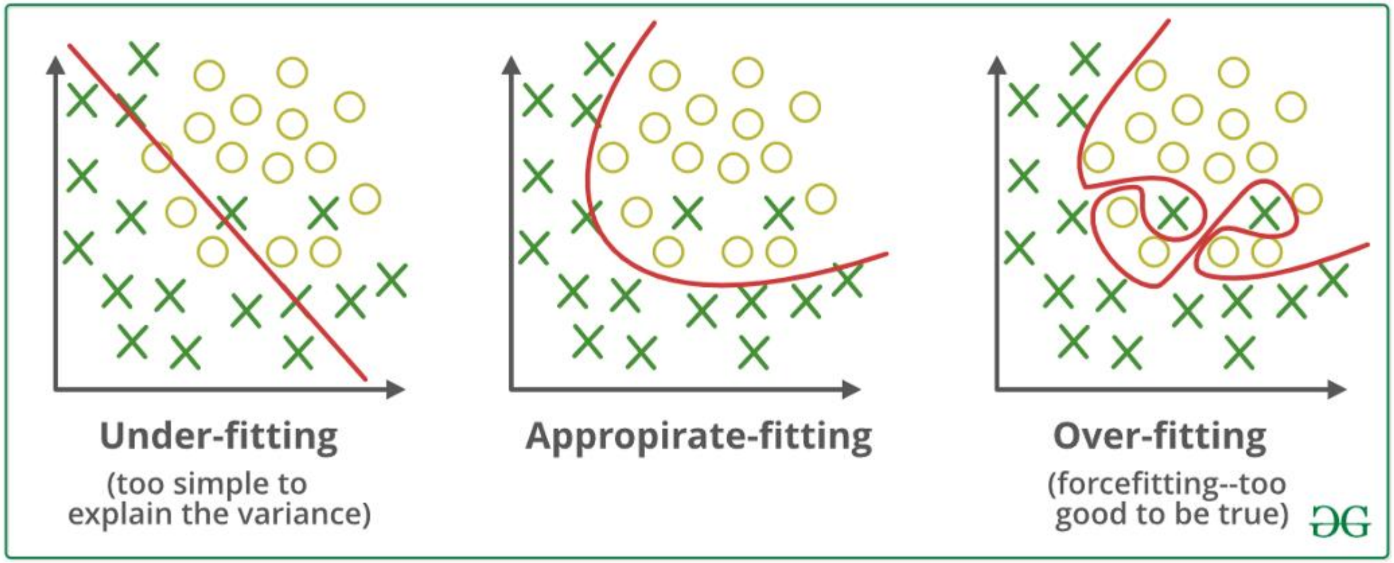


- **Training set**: a collection of instances from which a classifier is induced/trained
- **Test Set**: A collection of instances which were never used for learning the classifier
  - For measuring the performance of the learnt classifier

Generalisation

- We learn a classifier/predictor/model from the training data
- But performing well on training data is **NOT** enough!
- Important to evaluate the performance on the **test (unseen) data – generalisation** 评估测试（看不见的）数据的性能很重要–泛化

- If too biased to the training data, this may cause *overfitting*: too good on the training data, but poor on test data

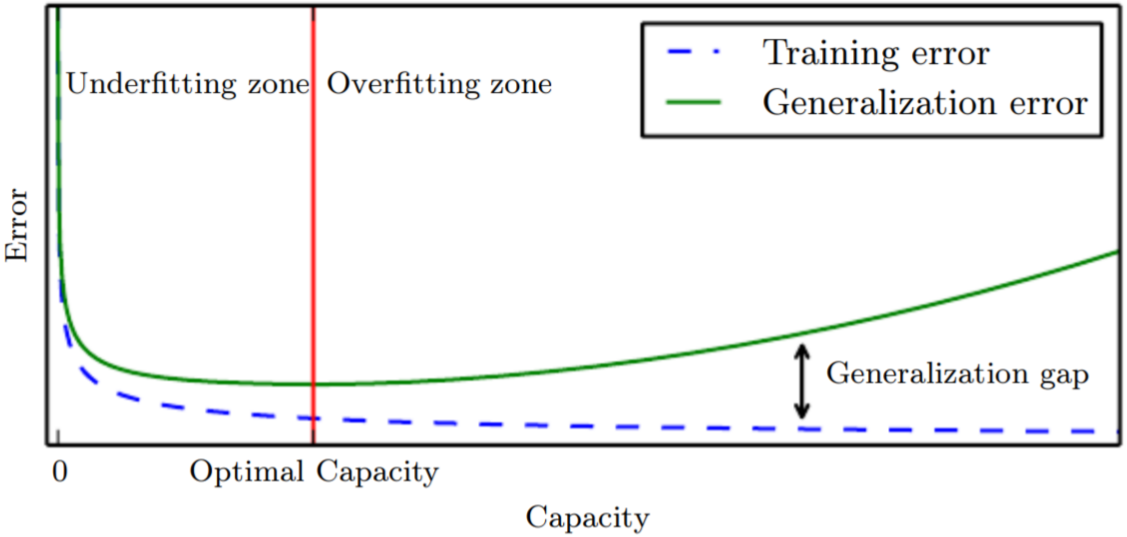

Under-fitting
(too simple to explain the variance)
Appropirate-fitting
Over-fitting
(forcefitting--too good to be true)

https://towardsdatascience.com/underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6fe4a8a49dbf

Generalisation

- Why? Our training data nearly always has some "signal" and some "noise".
- Learning **too** well means capturing the "noise"!

- E.g. one COMP307 student in 2020 is 2m tall, and gets an A+
  - Overfitted AI algorithm: "Students over 2m tall always get an A+!"
  - Well-fitted AI algorithm: doesn't consider height at all.



Summary

- Basic concepts of machine learning
- Categories of machine learning
- Common machine learning tasks
- Main machine learning paradigms/approaches
- Training set vs test set (vs validation set)
- Generalisation

- **Next lecture**: 3-K Techniques
- Suggested reading: online materials and sections 20.4 (2nd edi- tion) or sections 18.8 and 18.4 (3rd edition)