

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts

core definitions
attributes

Resources

Books
Tools

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge Technologies

<https://powcoder.com>

Sarah Erfani and Karin Verspoor, CIS

Add WeChat powcoder

Semester 2, 2017



THE UNIVERSITY OF
MELBOURNE

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

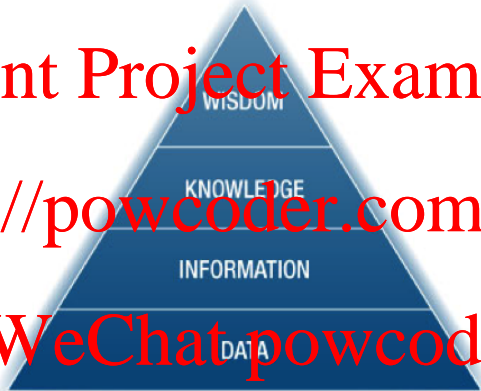
Books

Tools

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



[http://www.innovation.gov.au/Science/PMSEIC/Documents/
DataForScience.pdf](http://www.innovation.gov.au/Science/PMSEIC/Documents/DataForScience.pdf)

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

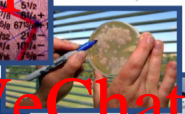
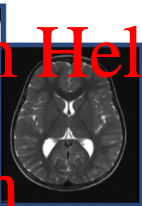
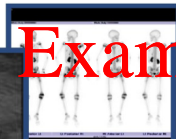
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



| | | | |
|------|------|-----|-----|
| 632 | 44% | 43 | 40% |
| 54 | 13 | 12 | 12 |
| 35 | 7 | 7 | 7 |
| 647 | 25% | 21% | 21% |
| 172 | 8% | 7% | 7% |
| 1906 | 21% | 25% | 25% |
| 35 | 51% | 49% | 5 |
| 1103 | 31% | 31% | 31% |
| 125 | 5% | 5% | 5% |
| 39 | 61% | 51% | 61% |
| 2302 | 681% | 65% | 67% |
| 2944 | 21% | 2% | 2% |
| 92 | 10% | 10% | 10% |
| 1019 | 77% | 7% | 7% |



Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

Information interpreted with respect to a user's context to extend human understanding in a given area.

<https://powcoder.com>

... In the context of data, perhaps:

Increasing insight into data, based on a user's information needs in a given context.

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Techniques

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

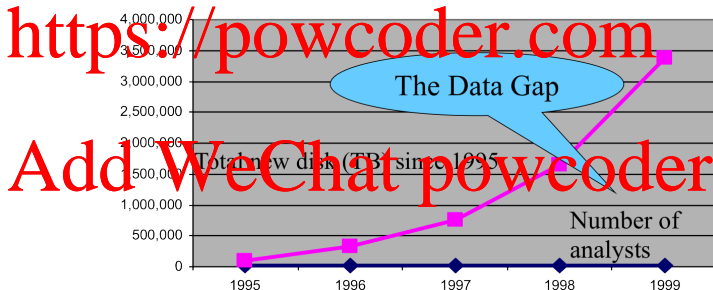
Tackling the challenge of knowledge management and discovery at a massive scale

- Database modelling and integration has long been a focus of Information Technology research and development. Classic example being the application of RDBMs for commercial apps.
- A major and accelerating trend is the focus of data integration from business and enterprise applications to scientific and personal applications.
- Exponential growth of data with the spread of the Internet Web and the multitudes of automatic data generation and collection devices.

This trend is expected to continue in the foreseeable future.

Importance of Problem

- Current computational methods cannot handle magnitude and dimensionality of the data
- Decision makers and Scientists need techniques to help form hypotheses and make evidence based decisions



Tools are required to integrate, distill, and make sense of data.

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Extracting

- implicit,
- *previously unknown*,
- potentially useful

information from data

- Needed: programs that detect patterns and regularities in the data
- Strong patterns → good predictions
 - Problem 1: most patterns are not interesting
 - Problem 2: patterns may be inexact (or spurious)
 - Problem 3: data may be garbled or missing

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

Arthur Samuel (1959)

<https://powcoder.com>

- "Field of study that gives computers the ability to learn without being explicitly programmed"

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Assignment Project Exam Help

Arthur Samuel (1959)

- “Field of study that gives computers the ability to learn without being explicitly programmed”

<https://powcoder.com>

Tom Mitchell (1999)

- “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 .”

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

Algorithms for acquiring structural descriptions from examples

- Structural descriptions represent patterns explicitly
- Can be used to predict outcome in new situation
- Can be used to understand and *explain* how prediction is derived (may be even more important)

Methods originate from artificial intelligence, statistics, and research on databases

<https://powcoder.com>

Add WeChat powcoder

Dictionary definitions of “learning”:

- To get knowledge of by study, experience, or being taught
- To become aware of information or from observation
- To commit to memory
- To be informed of, as certain, to receive instruction

→ Difficult to measure; Trivial for computers

Operational definition:

- Things learn when they change their behaviour in a way that makes them perform better in the future.
- Does learning imply intention?

Supervised learning

- *Teach* the computer how to do something (*by example*), then let it use its new-found knowledge to do it
- Labeled data: for given inputs, provide the expected output (“the answer”)
- Infer a function mapping from inputs to outputs

Unsupervised learning

- Let the computer *learn how to do something*
- Determine structure and patterns in data
- Unlabeled data: Don’t give the computer “the answer”

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

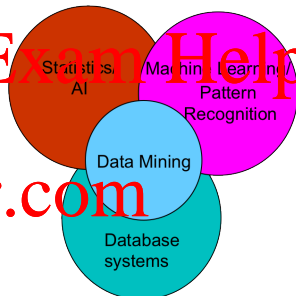
The distinctions between Data Mining and Machine Learning are not cut-and-dried. Data mining is primarily about discovering something hidden in your data, that you did not know before, as “new” as possible.

Knowledge obtained from data.

Machine learning emphasises algorithms used to generalise existing knowledge to new data, as accurately as possible.

Techniques used to learn from data.

Data mining applications typically use a lot of machine learning techniques. For example a pattern in a data set that is useful for generalisation might represent new knowledge.



<https://powcoder.com>

Add WeChat powcoder

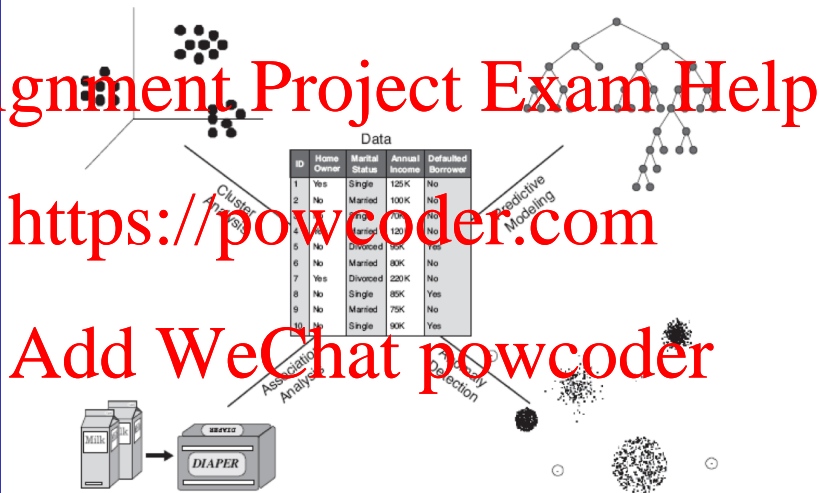


Figure 1.3. Four of the core data mining tasks.

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and Problems

ML applications

Types of ML
Examples

ML concepts

core definitions
attributes

Resources

Books
Tools

Supervised learning

- Classification
predicting a discrete class
- Regression
predicting a numeric quantity

Unsupervised learning

- Association
detecting associations between features
- Information organisation, Clustering
grouping similar instances into clusters
- Reinforcement learning
- Recommender systems
- Anomaly/outlier detection

Example: Supervised Learning (Regression)

Introduction to
Data Mining and
Machine Learning

COMP90049

Knowledge

Techniques

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

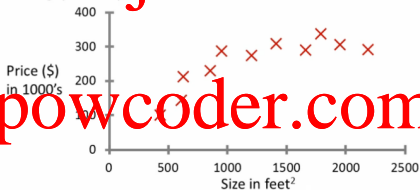
Resources

Books

Tools

Can we predict housing prices?

Housing price prediction.



<https://powcoder.com>

Add WeChat powcoder

A friend has a house which is 750 square feet – how much can he expect to get?

(draw a straight line vs. fit a curve)

Example: Unsupervised Learning

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and Problems

ML applications

Types of ML
Examples

ML concepts
core definitions
attributes

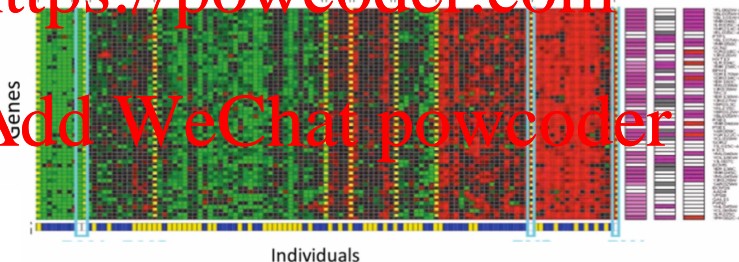
Resources

Books
Tools

- Given gene expression data for individuals, cluster based on expression profiles
- Group newspaper articles into cohesive groups
- Credit card fraud
- Network intrusion behaviour

<https://powcoder.com>

Add WeChat powcoder



Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Techniques

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

- The input to a machine learning system consists of:

- *Instances*: the individual, independent examples of a concept
also known as exemplars

- *Attributes*: measuring aspects of an instance
also known as features

- *Concepts*: things that we aim to learn

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Example: Supervised Learning (Classification)

Introduction to
Data Mining and
Machine Learning

COMP90049

Knowledge
Techniques

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

| Outlook | Temperature | Humidity | Windy | Play |
|----------|-------------|----------|-------|------|
| sunny | hot | high | FALSE | no |
| sunny | hot | high | TRUE | no |
| overcast | hot | high | FALSE | yes |
| rainy | mild | high | FALSE | yes |
| rainy | cool | normal | FALSE | yes |
| rainy | cool | normal | TRUE | no |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Given information about current weather conditions and the forecast, can we determine whether we will go out to play?

Assignment Project Exam Help

<https://powcoder.com>

Add: WeChat : powcoder:

| Outlook | Temperature | Humidity | Windy | Play |
|----------|-------------|----------|-------|-----------------|
| sunny | hot | high | FALSE | no ₁ |
| sunny | hot | high | TRUE | no ₂ |
| overcast | hot | high | FALSE | yes |
| rainy | mild | high | FALSE | yes |
| rainy | cool | normal | FALSE | yes |
| rainy | cool | normal | TRUE | no |

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat : powcoder:

| Outlook | Temperature | Humidity | Windy | Play |
|----------|-------------|----------|-------|------|
| sunny | hot | high | FALSE | no |
| sunny | hot | high | TRUE | no |
| overcast | hot | high | FALSE | yes |
| rainy | mild | high | FALSE | yes |
| rainy | cool | normal | FALSE | yes |
| rainy | cool | normal | TRUE | no |

Assignment Project Exam Help

<https://powcoder.com>

Add: WeChat : powcoder

| Outlook | Temperature | Humidity | Windy | Play |
|----------|-------------|----------|-------|------|
| sunny | hot | high | FALSE | no |
| sunny | hot | high | TRUE | no |
| overcast | hot | high | FALSE | yes |
| rainy | mild | high | FALSE | yes |
| rainy | cool | normal | FALSE | yes |
| rainy | cool | normal | TRUE | no |

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Assignment Project Exam Help

Each instance is described by a fixed feature vector

Possible attribute types (levels of measurement):

- nominal
- ordinal
- continuous

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Techniques

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

- Values are distinct symbols (e.g. {sunny,overcast,rainy})
 - values themselves serve only as labels or names

- Also called *categorical, enumerated, or discrete* (NB: “enumerated” and “discrete” imply an order which tends not to exist)

- Special case: dichotomy (“boolean” attribute)

- No relation is implied among nominal values (no ordering or distance measure), and only equality tests can be performed

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

- An explicit order is imposed on the values (e.g. {hot,mild,cool} where $\text{hot} > \text{mild} > \text{cool}$)

- No distance between values defined, addition and subtraction don't make sense

- Example rule: $\text{temperature} < \text{hot} \rightarrow \text{play} = \text{yes}$

- Distinction between nominal and ordinal not always clear (e.g. our Look)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

- Continuous features are real-valued with a well-defined zero point and no explicit upper bound

- Also called *numeric*

- Example: attribute distance

Distance between an object and itself is zero

- All mathematical operations are allowed

Introduction to Data Mining and Machine Learning

COMP90049

Knowledge

Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

How might you approach data mining the Weather dataset?

- Methods

- Using Supervised methods?
- Using Unsupervised methods?

- Attributes

- Are there regularities among the attributes?
- Are there different ways you could make use of the attributes (e.g. different combinations? different thresholds? ?)

<https://powcoder.com>

Add WeChat powcoder

Introduction to Data Mining and Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books

Tools

Assignment Project Exam Help

Introduction to Data Mining

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2006. Addison
Wesley

<http://www-users.cs.umn.edu/~kumar/dmbook/index.php>

Data Mining: Practical Machine Learning Tools and Techniques

Ian Witten, Eibe Frank, Mark Hall

<http://www.cs.waikato.ac.nz/ml/weka/book.html>

Add WeChat powcoder

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Assignment Project Exam Help

WEKA Toolkit

<http://www.cs.waikato.ac.nz/ml/weka/index.html>

List of more specific tools

<http://www-users.cs.umn.edu/~kumar/dmbook/resources.htm>

Data sets

UC Irvine Machine Learning Data Repository

<http://archive.cs.uci.edu/ml/datasets.html>

<https://powcoder.com>

Add WeChat powcoder