COMP2610 / COMP6261 Information Theory

Lecture 1: Introduction

Robert C. Williamson

Research School of Computer Science
The Australian National University

July 23, 2018

What is the world made of?

- Ancient times: **Matter** — atoms

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)
- **Economics** (price, markets, the economics of information)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)
- **Economics** (price, markets, the economics of information)
- **Sociology** (media, social networks)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)
- **Economics** (price, markets, the economics of information)
- **Sociology** (media, social networks)
- **Philosophy** (ontology, epistemology, morality)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)

- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)

- **Economics** (price, markets, the economics of information)
- **Sociology** (media, social networks)
- **Philosophy** (ontology, epistemology, morality)
- **Engineering** (your telephone for example)

What is the world made of?

- Ancient times: **Matter** — atoms
- 20th Century: **Energy** — mass=energy
- 21st Century: **Information** — ????

Information underpins

- **Physics** (energy needs of computing limited by cost of *erasing* information)
- **Chemistry** (pattern recognition by molecules)
- **Biology** (genetic code)
- **Immunology** (pattern recognition of self from non-self)
- **Economics** (price, markets, the economics of information)
- **Sociology** (media, social networks)
- **Philosophy** (ontology, epistemology, morality)
- **Engineering** (your telephone for example)
- **Computing** (What is that computers do? *They process information*)

References for the curious . . . for interest only!

1. Tom Siegfried, *The Bit and the Pendulum: From Quantum Computing to M Theory-The New Physics of Information*, Wiley 2000
2. Giles Brassard, Is information the Key?, *Nature Physics* 1, 1-4, October 2006
3. John Archibald Wheeler, Information, Physics, Quantum: The Search for Links, in *Proceedings of the 3rd International Symposium on the Foundations of Quantum Mechanics,* Tokyo, (1989)
4. John Archibald Wheeler with Kenneth Ford, *Geons, Black Holes, and Quantum Foam: A Life in Physics*, W.W. Norton and Company, 1998; Chapter 15 "It from Bit"
5. Paul Davies and Niels Henrik Gregersen, *Information and the Nature of Reality*, Cambridge University press 2010
6. Andreas Wagner, From bit to it: how complex metabolic network transforms information into living matter, *BMC Systems Biology*, 1(33), 2007
7. Hector Zenil (Ed.), *A computable universe: understanding and exploring nature as computation*, World Scientific (2013)
8. Rolf Landauer, Uncertainty principle and minimal energy dissipation in the computer, *International Journal of Theoretical Physics* 21(3/4), 283-297, (1982)
9. Rolf Landauer, The physical nature of information, *Physics Letters* A, 217, 188-193 (1996)
10. Antonie Berut et al., Experimental verification of Landauer's principle linking information and thermodynamics, *Nature* 483, 187-190, (8 March 2012)
11. Juan M.R. Parrondo, Jordan M. Horowitz and Takahiro Sagawa, Thermodynamics of Information, *Nature Physics*, 11, 131-139, (February 2015)
12. Jean-Marie Lehn, Perspectives in Supramolecular Chemistry —– From Molecular Recognition towards Molecular Information Processing and Self-Organization, *Angewandte Chemie International Edition in English*, 29(11), 1304–1319, (November 1990)
13. Jean-Marie Lehn, *Supramolecular chemistry – scope and perspectives – molecules – supermolecules – molecular devices*, Nobel Prize Lecture, (8 December 1987)
14. John Maynard Smith, The concept of information in biology, *Philosophy of Science* 67(2), 177-194 (2000)
15. Ladislav Kovac, Information and knowledge in biology: time for reappraisal, *Plant Signalling and behaviour* 2(2), 65-73 (2007)
16. David Easley and Jon Kleinberg, *Networks, crowds and markets: reasoning about a highly connected world*, Cambridge University Press (2010).
17. Friedrich A. Hayek, The use of knowledge in society, *The American Economic Review*, 35(4), 519-530 (1945)
18. George J. Stigler, The Economics of Information, *The Journal of Political Economy* 69(3), 213-225 (1961)
19. Joseph E. Stiglitz, *Information and the change in the paradigm in economics*, Nobel Prize Lecture 8 (December 2001)
20. Warwick Anderson and Ian R. Mackay, Fashioning the immunological self: the biological individuality of F. Macfarlane Burnet, *Journal of the History of Biology*, 47, 147–175, (2014)

What Is Information? (1)

According to a dictionary definition, information can mean

1. Facts provided or learned about something or someone:
   a vital piece of information.

2. What is conveyed or represented by a particular arrangement or sequence of things:
   genetically transmitted information.

What Is Information? (1)

According to a dictionary definition, information can mean

1. Facts provided or learned about something or someone:
   a vital piece of information.
2. What is conveyed or represented by a particular arrangement or sequence of things:
   genetically transmitted information.

Important!

- Usually unhelpful to ask “What is?” questions! — “essentialism”.

What Is Information? (1)

According to a dictionary definition, information can mean

1. Facts provided or learned about something or someone:
   a vital piece of information.

2. What is conveyed or represented by a particular arrangement or sequence of things:
   genetically transmitted information.

Important!

- Usually unhelpful to ask "What is?" questions! — "essentialism".
- Better to ask what happens to it? "Grothendieck's Relative method"

What is Information? (2)

In this course: information in the context of *communication* (includes information storage).

- Explicitly include uncertainty — indeed, rather than deriving information from probability theory, one can start with information and derive probability theory from that!

- Claude Shannon (1948): "Amount of unexpected data a message contains"
  - A theory of information **transmission**

What is Information? (3)

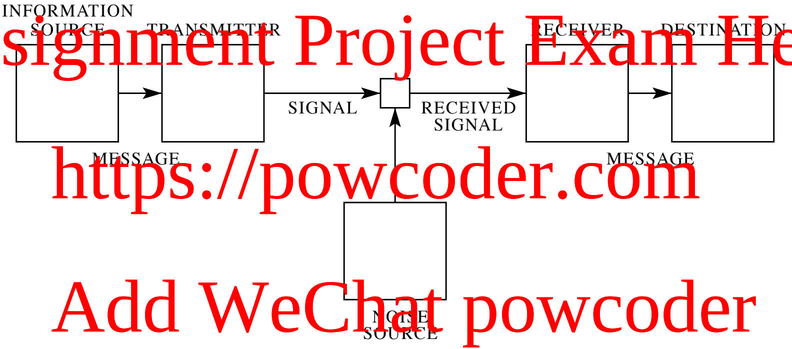


Fig. 1 — Schematic diagram of a general communication system.

From Claude Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal* (1948).

What Is Information? (4)

Information is a message that is *uncertain* to receivers:

- If we receive something that we already knew with absolute certainty then it is non-informative

- Uncertainty is crucial in measuring information content

- We will deal with uncertainty using probability theory

What Is Information? (4)

Information is a message that is *uncertain* to receivers:

- If we receive something that we already knew with absolute certainty then it is non-informative

- Uncertainty is crucial in measuring information content

- We will deal with uncertainty using probability theory

Information Theory

Information theory is the study of the fundamental *limits* and *potential* of the **representation** and **transmission** of information.

Examples

Example 1: What Number Am I Thinking of?

- I have in mind a number that is between 1 and 20
- You are allowed to ask me one question at a time
- I can only answer yes/no
- Your goal is to figure out the number **as quickly as possible**
- What strategy would you follow?

Example 1: What Number Am I Thinking of?

- I have in mind a number that is between 1 and 20
- You are allowed to ask me one question at a time
- I can only answer yes/no
- Your goal is to figure out the number **as quickly as possible**
- What strategy would you follow?

Your strategy + my answers = a code for each number

Some variants:

- What if you knew I never chose prime numbers?
- What if you knew I was twice as likely to pick numbers more than 10?
- What if you knew I only ever chose one of 7 or 13?

Example 2: How Much Is Information Worth?

Simplified Version of “Deal or No Deal”

$1000 Hidden in one of 16 cases.

- All equally likely to contain the prize

Example 2: How Much Is Information Worth?

Simplified Version of “Deal or No Deal”

$1000 Hidden in one of 16 cases.

- All equally likely to contain the prize

How much would you pay to know:

1. Exactly which case contains the money?
2. Whether the case holding the money is numbered less than 8?
3. ... is less than 12?
4. Which range out of 0–3, 4–7, 8–11, or 12–15 the money case is in?

Example 2: How Much Is Information Worth?

Simplified Version of “Deal or No Deal”

$1000 Hidden in one of 16 cases.

- All equally likely to contain the prize

How much would you pay to know:

1. Exactly which case contains the money?

2. Whether the case holding the money is numbered less than 8?
3. ... is less than 12?

4. Which range out of 0–3, 4–7, 8–11, or 12–15 the money case is in?

Key Question:

- Can we use these ideas to *quantify* information?

Example 3: Redundancy and Compression

Cn y rd ths sntnc wtht ny vwls?

Example 3: Redundancy and Compression

Cn y rd ths sntnc wtht ny vwls?

Can you read this sentence without any vowels?

Written English (and other languages) has much redundancy:

- Approximately 1 bit of information per letter

- Naïvely there should be almost 5 bits per letter

(For the moment think of “bit” as “number of yes/no questions”)

Example 3: Redundancy and Compression

Cn y rd ths sntnc wtht ny vwls?

Can you read this sentence without any vowels?

Written English (and other languages) has much redundancy:

- Approximately 1 bit of information per letter

- Naïvely there should be almost 5 bits per letter

(For the moment think of “bit” as “number of yes/no questions”)

Key Question:

- How much redundancy can we *safely* remove?
  (Note: “rd” could be “read”, “red”, “road”, etc.)

Example 4: Error Correction

Hmauns hvae the aitliby to cerroct for eorrrs in txet and iegmas.

Key Question:

- How much noise is it possible to correct for and how?

A Summary of the History of Information Theory

1920s : Nyquist & Hartley at Bell Labs

1940 : Turing and Good at Bletchley Park (WWII)

1942 : Hedy Lamarr and George Antheil

1948 : Claude Shannon: "A Mathematical Theory of Communication"

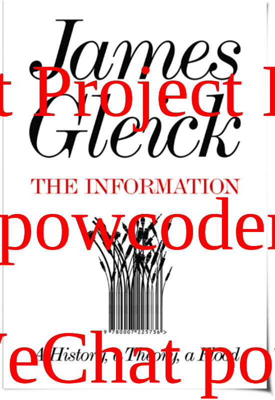
1951 : Huffman Coding

1958 : Peter Elias: "Two Famous Papers"

1970 : "Coding is Dead"

1970- : Revival with advent of digital computing
CDs, DVDs, MP3s, Digital TV, Mobiles, Internet, Deep-space comms (Voyager), ...

More on the History of Information Theory



&

Information Theory and the Digital Age

by Aftab, Cheung, Kim, Thakkar, and Yeddanapudi.

`http://web.mit.edu/6.933/www/Fall2001/Shannon2.pdf`

1. Information and the Nature of the Universe
2. A Brief History
3. Course Overview
4. Logistics and Expectations
5. What's Next

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information
- How can we make good guesses?
  - Probabilistic Inference
  - Bayes Theorem

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information
- How can we make good guesses?
  - Probabilistic Inference
  - Bayes Theorem
- How much redundancy can we safely remove?
  - Compression
  - Source Coding Theorem, Kraft Inequality
  - Block, Huffman, and Lempev-Ziv Coding

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information
- How can we make good guesses?
  - Probabilistic Inference
  - Bayes Theorem
- How much redundancy can we safely remove?
  - Compression
  - Source Coding Theorem, Kraft Inequality
  - Block, Huffman, and Lempev-Ziv Coding
- How much noise can we correct and how?
  - Noisy-Channel Coding
  - Repetition Codes, Hamming Codes

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information
- How can we make good guesses?
  - Probabilistic Inference
  - Bayes Theorem
- How much redundancy can we safely remove?
  - Compression
  - Source Coding Theorem, Kraft Inequality
  - Block, Huffman, and Lempev-Ziv Coding
- How much noise can we correct and how?
  - Noisy-Channel Coding
  - Repetition Codes, Hamming Codes
- What is randomness?
  - Kolmogorov Complexity
  - Algorithmic Information Theory

Brief Overview of Course

- How can we quantify information?
  - Basic Definitions and Key Concepts
  - Probability, Entropy & Information
- How can we make good guesses?
  - Probabilistic Inference
  - Bayes Theorem
- How much redundancy can we safely remove?
  - Compression
  - Source Coding Theorem, Kraft Inequality
  - Block, Huffman, and Lempev-Ziv Coding
- How much noise can we correct and how?
  - Noisy-Channel Coding
  - Repetition Codes, Hamming Codes
- What is randomness? [Marcus Hutter]
  - Kolmogorov Complexity
  - Algorithmic Information Theory

COMP2610/COMP6261 (Information Theory)

We will study the fundamental limits and potential of the *representation* and *transmission* of information.

- Mathematical Foundations
- Probabilistic Inference

- Coding and Compression

- Communication
- Kolmogorov Complexity (Guest Lecture)

Learning Outcomes

From `https://wattlecourses.anu.edu.au/course/view.php?id=25550`:

1. Understand and apply **fundamental concepts** in information theory such as probability, entropy, information content and their inter-relationships
2. Understand the principles of **data compression**
3. Compute **entropy** and **mutual information** of random variables
4. Implement and analyse basic **coding** and **compression algorithms**
5. Understand the relationship of information theoretical principles and **Bayesian inference** in data modelling and pattern recognition
6. Understand some key **theorems** and **inequalities** that quantify essential limitations on compression, communication and inference
7. Know the basic concepts regarding **communications over noisy channels**

What Tools Will We Use?

- Elementary probability theory
  - "What's the probability of rolling an odd number using a fair die?"

What Tools Will We Use?

- Elementary probability theory
  - "What's the probability of rolling an odd number using a fair die?"

- Elementary linear algebra
  - "If $x = (1, 1, 0)$ and $y = (-2, 0, 1)$ what is $x \cdot y$ and $3x + 2y$?"

What Tools Will We Use?

- Elementary probability theory
  - "What's the probability of rolling an odd number using a fair die?"

- Elementary linear algebra
  - "If $x = (1, 1, 0)$ and $y = (-2, 0, 1)$ what is $x \cdot y$ and $3x + 2y$?"
- Basic programming skills
  - "Do you know your `for` loops from your `while` loops?"

What Tools Will We Use?

- Elementary probability theory
  - "What's the probability of rolling an odd number using a fair die?"
  - http://www.khanacademy.org/math/probability

- Elementary linear algebra
  - "If $x = (1, 1, 0)$ and $y = (-2, 0, 1)$ what is $x \cdot y$ and $3x + 2y$?"
  - http://www.khanacademy.org/math/linear-algebra
- Basic programming skills
  - "Do you know your `for` loops from your `while` loops?"

Outline

1. Information and the Nature of the Universe
2. A Brief History
3. Course Overview
4. Logistics and Expectations
5. What's Next

Course Overview

See Wattle site (authoritative)

- Lectures: 23 $\times$ 1 hour (two lectures per week); one public holiday

- By me, except one guest lecture by Marcus Hutter (Aside: about me).
- Tutorials: Starting week 2; schedule up shortly.

- Assignments: 3 (0% (optional), 20%, 20% each) (0% explained below)

- Final Exam (60%) Hurdle assessment. You have to pass the exam to pass the course. (New this year!)
- **Late Submission Policy: late submissions get zero marks — 100% penalty.**

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. "Assignment 0" is designed to help you check whether your background is sufficient.

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. “Assignment 0” is designed to help you check whether your background is sufficient.
- As per the expectations document: you are responsible for your learning. I am here to *assist*. I take this seriously (I wrote the first draft of the expectations doc!)

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. “Assignment 0” is designed to help you check whether your background is sufficient.
- As per the expectations document: you are responsible for your learning. I am here to *assist*. I take this seriously (I wrote the first draft of the expectations doc!)
- You are *not* obliged to attend any of the lectures or tutorials: you are adults. Not attending is a high risk strategy!

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. “Assignment 0” is designed to help you check whether your background is sufficient.
- As per the expectations document: you are responsible for your learning. I am here to *assist*. I take this seriously (I wrote the first draft of the expectations doc!)

- You are *not* obliged to attend any of the lectures or tutorials: you are adults. Not attending is a high risk strategy!

- The course closely follows the text. In principle, you can study that, do exercises, skip all lectures and tuts and get a HD.

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. “Assignment 0” is designed to help you check whether your background is sufficient.
- As per the expectations document: you are responsible for your learning. I am here to *assist*. I take this seriously (I wrote the first draft of the expectations doc!)
- You are *not* obliged to attend any of the lectures or tutorials: you are adults. Not attending is a high risk strategy!
- The course closely follows the text. In principle, you can study that, do exercises, skip all lectures and tuts and get a HD.
- If you do come to lectures, please come on time, pay attention, and put your telephone on silent. (Basic politeness)

Expectations

See the newly published expectations document:

https://wattlecourses.anu.edu.au/pluginfile.php/1760092/course/section/423322/Learning%20expectations.pdf

Key points:

- You are expected to have familiarity and ability with elementary probability theory. "Assignment 0" is designed to help you check whether your background is sufficient.
- As per the expectations document: you are responsible for your learning. I am here to *assist*. I take this seriously (I wrote the first draft of the expectations doc!)

- You are *not* obliged to attend any of the lectures or tutorials: you are adults. Not attending is a high risk strategy!

- The course closely follows the text. In principle, you can study that, do exercises, skip all lectures and tuts and get a HD.
- If you do come to lectures, please come on time, pay attention, and put your telephone on silent. (Basic politeness)
- Learning mathematical material is hard and cannot be delegated or outsourced. "There is no royal road to geometry." Don't kid yourself!

Tutorials

- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures

Tutorials

- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures
- You are expected to have tried the exercises beforehand. Do not think you can just turn up an watch. Or get someone else to do it for you.

Tutorials

- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures
- You are expected to have tried the exercises beforehand. Do not think you can just turn up an watch. Or get someone else to do it for you.
- You cannot learn maths by watching someone else do it. Just like riding a bike; cooking; programming; piano; everything!

Tutorials

- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures
- You are expected to have tried the exercises beforehand. Do not think you can just turn up an watch. Or get someone else to do it for you.
- You cannot learn maths by watching someone else do it. Just like riding a bike; cooking; programming; piano; everything!
- You will get far more from a tutorial by trying the questions; failing; and *then* seeing what you should have done.

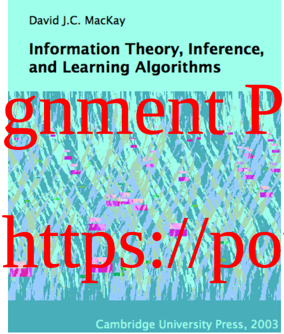
Tutorials

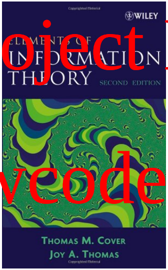
- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures
- You are expected to have tried the exercises beforehand. Do not think you can just turn up an watch. Or get someone else to do it for you.
- You cannot learn maths by watching someone else do it. Just like riding a bike; cooking; programming; piano; everything!
- You will get far more from a tutorial by trying the questions; failing; and *then* seeing what you should have done.
- This is not merely my opinion. There is extensive evidence! Anders Ericsson and Robert Pool, *Peak: Secrets from the New Science of Expertise*, Houghton Mifflin Harcourt, 2016.

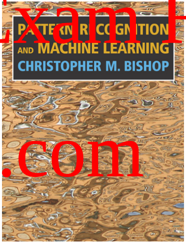
Tutorials

- Problem sets of exercises will be provided for each tutorial
- These will review material covered in previous lectures
- You are expected to have tried the exercises beforehand. Do not think you can just turn up an watch. Or get someone else to do it for you.
- You cannot learn maths by watching someone else do it. Just like riding a bike; cooking; programming; piano; everything!
- You will get far more from a tutorial by trying the questions; failing; and *then* seeing what you should have done.
- This is not merely my opinion. There is extensive evidence! Anders Ericsson and Robert Pool, *Peak: Secrets from the New Science of Expertise*, Houghton Mifflin Harcourt, 2016.
- In a nutshell: The secret of success is *deliberate practice.*

Textbook



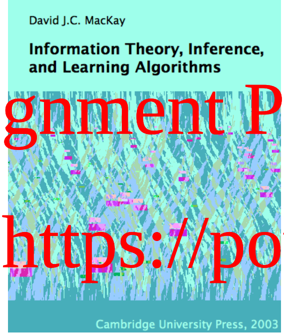


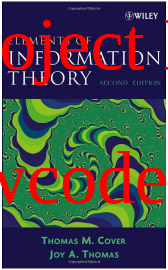


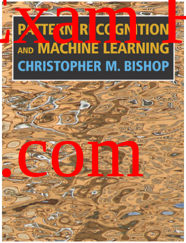
Mackay (ITILA, 2006) available online:
`http://www.inference.phy.cam.ac.uk/mackay/itila`

- Note copyright rules: e.g. copying the whole book onto paper is not permitted.
- We will follow a different chapter order to that given in the book

Textbook







Mackay (ITILA, 2006) available online:
http://www.inference.phy.cam.ac.uk/mackay/itila

- Note copyright rules: e.g. copying the whole book onto paper is not permitted.
- We will follow a different chapter order to that given in the book

For an alternative take – David MacKay's Lectures:
http://www.inference.phy.cam.ac.uk/itprnn_lectures/

Consultation & Other Issues

Consultation:

- Best way to contact the course lecturers and tutors is via **email**

  comp2610@anu.edu.au

- If you **really** need to meet in person, send an email request first

- Email response times may vary but consider 1 day as a fast reply and up to three days as a normal response time

- **Technical questions**: encouraged to post on Wattle's public forum

- Request for clarifying assignment: **must be** posted on Wattle

What's Next?

- If you are not comfortable about your probability and algebra skills, start today on improving them

- Get a copy of the text and start purusing it

- Sign up to a tutorial (will open tomorrow, time announced tomorrow)