COMP2610 / COMP6261 Information Theory
Lecture 1: Introduction

Thushara Abhayapala

Audio & Acoustic Signal Processing Group
School of Engineering,
The Australian National University,
Canberra, Australia.

25 July 2023

Overview

1. Course Outline - Mechanics
2. Information and the Nature of the Universe
3. Examples
4. Brief History

Course Information and Communications

We use ANU Wattle site for all course communications
`https://wattlecourses.anu.edu.au/course/view.php?id=41135`

- Check Course Outline document
- It is your responsibility to check the webpage everyday for course information and announcements.
- Use ‘Course Discussion Forum’ for questions.

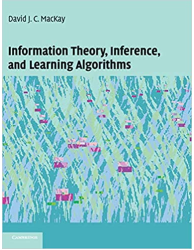
Teaching Team

- Thushara Abhayapala (Convenor)
- Manish Kumar (Manish.Kumar@anu.edu.au)
- Angela Zhang (Yile.Zhang@anu.edu.au)
- Naisheng (Nick) Liang (u6356745@anu.edu.au)
- Zhifeng Tang (Zhifeng.Tang@anu.edu.au)

Pre-Requisite and Self-administered Quiz

- No formal per-requisite
- Working familiarity with elementary probability theory - Understanding of
  - notions of probability, conditional probability,
  - expectation, other moments,
  - distribution functions, density functions, joint distributions.
- The 'Teaching Team' will very quickly remind you of these concepts in the lectures. But we are not teaching them from scratch.
- Self-administered Quiz - Now available on Wattle under 'Assignments'. (Not a formal assessment)

Text Book



David MacKay, "Information Theory, Inference, and Learning Algorithms", Cambridge University press, (primary text; available http://www.inference.org.uk/mackay/itila)

Additional reading (Available in ANU Library):

- "Elements of Information Theory" by Cover and Thomas, 2nd Edition, New York, Wiley, 2006.
- "Pattern Recognition and Machine Learning," by Christopher M. Bishop

Lectures and Tutorials

- Two Lectures per week
- Tutorials - Choose one from 7 repeat tute slots (Sign-up via MytimeTable). All tutes are on-campus only.
- Tutorials start from Week1 !
- Problem sets will be provided for each tutorial. These will review material covered in previous lectures except in week 1.
- You are expected to have tried the exercises beforehand. We will run tutorials in Workshop style.
- Do not think you can just turn up and 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.
- In a nutshell: The secret of success is deliberate practice.

Assessment

- COMP2610 and COMP6261 share some assessment. You will have to do a different subset of questions in the assignments and the exam depending upon which course you are enrolled in.
- There are FOUR components to the assessment for this course:
  - Assignment 1 10%
  - Assignment 2 20%
  - Assignment 3 20%
  - Final Exam 50% **Hurdle component, min score required is 40% of the exam**
- **Late Submission Policy:** A late submission attracts a penalty of 5% per working day as per ANU Policy until a week from the due date. We will provide solutions to the assignment after a week from the due date and if you submit after that time you get zero marks (100% penalty). Extensions will be considered according to the ANU Policy

Expectations

- You are expected to have familiarity and ability with elementary probability theory. The take-home quiz is designed to help you check whether your background is sufficient.
- You are responsible for your learning. We (Teaching Team) here to assist. We take this seriously.
- You are not obliged to attend any of the lectures or tutorials. 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 tutes and get a HD.
- Learning mathematical material is hard and cannot be delegated or outsourced. " There is no royal road to geometry." Don't kid yourself!

Consultation & Other Issues

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

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)
- **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*)

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!
- Better to ask what happens to it?

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**
  - Does not consider the **meaning** of the message . . . This is (arguably) in the eye of the beholder.
  - “The meaning of information is given by the processes that interpret it”

    N. Katherine Hayles, Cognition Everywhere: The Rise of Cognitive Nonconsciousness and the Costs of Consciousness, *New Literary History* 45, 199–220 (2014)

What is Information? (3)

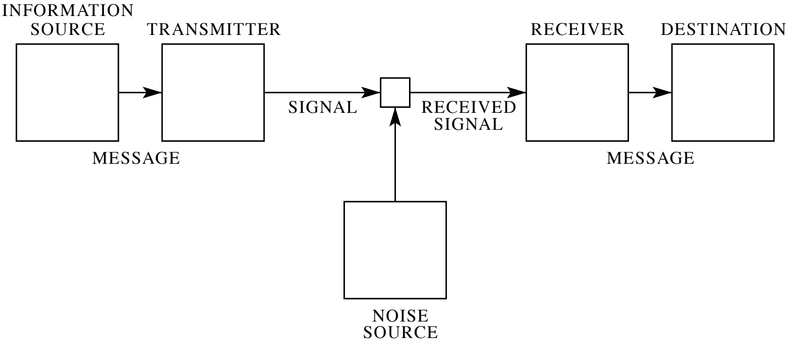


Fig. 1 — Schematic diagram of a general communication system.

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

Definition (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?

$$\text{Your strategy} \quad + \quad \text{my answers} = \text{a code for each number}$$

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:

- Exactly which case contains the money?
- Whether the case holding the money is numbered less than 8?
- ... is less than 12?
- 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:

- Exactly which case contains the money?
- Whether the case holding the money is numbered less than 8?
- ... is less than 12?
- 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
- Naively 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
- Naively 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), ...

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

Leaning Outcomes

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 and be able to implement classical compression schemes by hand on toy problems.
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?"
  - http://www.khanacademy.org/math/probability
  - Also see resources pointed to on Wattle
- 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?"

What's next

1. If you are not comfortable about your probability and algebra skills, start today on improving them
2. Get a copy of the text and start perusing it
3. Do the Self Assessment Quiz
4. Attend tutorials in week 1