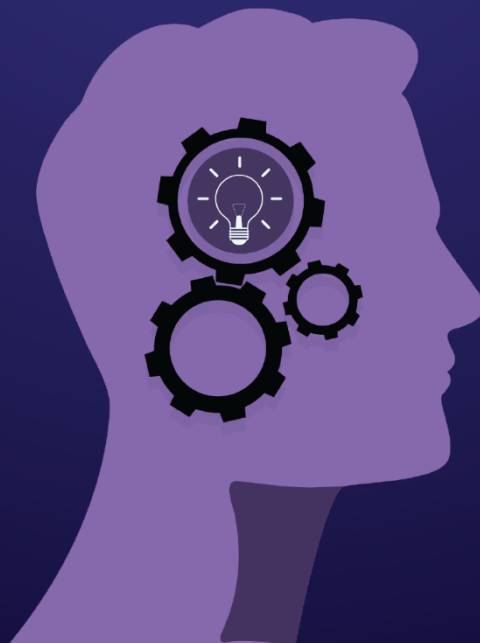




Information theory





تعليمات قبل ما تقرا الملخص :

- اللون الاسود هو الشرح
- الكلام اللي في الاسلايد مكتوب **باسود بولد**
- التعريفات باللون **الاصفر**
- الكلام المهم باللون **الاحمر**
- اللي بلون **ازرق** معلومات مش مهمة موجوده في المحاضرة بس حطيتها في الملخص تحسب لو جت في الامتحان
- هتلاقي في بداية الملخص تمهيد للمحاضرة فيه معلومات قديمة درسناها هتساعدك على فهم المحاضرة اكثر
- تقدر تروح لاخر صفحه في الملخص هتلاقي فيها ملخص المحاضرة كلها في صفحة واحده بدون شرح

مقدمة للمحاضرة :

- لازم تكون عارف ان ال information theory علم مبني على الرياضيات
- و اتعمل كجزء من ال communication و هو العلم اللي بيدرس طريقة نقل ال data من جهاز للتاني
- ولكن بعد ظهور ال AI اتاخدت معادلات ال Information theory وبقت بتستخدم في ال AI كمان

🔍 ايه هو ال data compression ؟

- هو تقليل لحجم ال data سواء كانت file او image او غيره
- وفيه منه نوعين : lossy compression و lossless compression

فلو ال data كانت صورة مثلا فهيبقي الوضع كالاتي :

- ال lossy فكرته انك بتقلل حجم الصورة اللي عندك مع فقد في " المعلومات " اللي في الصورة
- ال lossless فكرته انك بتقلل حجم ال data بس بدون فقد يذكر في " المعلومات " اللي في الصورة

🔍 خصائص اللوغاريتمات :

| Property Name | Property |
|--------------------------------|---|
| Log of 1 | $\log_a 1 = 0$ |
| Log of the same number as base | $\log_a a = 1$ |
| Product Rule | $\log_a (mn) = \log_a m + \log_a n$ |
| Quotient Rule | $\log_a \left(\frac{m}{n}\right) = \log_a m - \log_a n$ |
| Power Rule | $\log_a m^n = n \log_a m$ |
| Change of Base Rule | $\log_b a = \frac{\log_c a}{\log_c b}$ (OR) $\log_b a \cdot \log_c b = \log_c a$ |

1. What is information theory?

The central problem in information theory is **efficient** and **reliable** transmission of data from a source to a destination.

الكلمتين اللتي بالاحمر مهمين وييجوا في ال mcq

العلم ده هدفه حاجتين :

لـ اولا : انه يحدد ايه افضل " كفاءة " لنقل الداتا والمقصود نحدد افضل algorithm ممكن نقدر نضغط بيه الداتا .

معنى efficient بيخلصه السؤال ده

How much can data be compressed? (The best possible compression that can be achieved)

والاجابة على السؤال ده طلعت لنا نظرية اسمها ال Source coding لانها عبارة عن algorithm بيطبق عند ال source واللي هو المصدر اللتي ببيعت الداتا , وتعرفها هو

Source coding theorem: The **maximum** rate at which data can be **compressed losslessly** is the entropy rate of the source.

لـ الهدف التاني : انه ينقل الداتا او المعلومات بدون ما يحصل ليها تشويش او اختراق , يعني يكون " واثق " ان الداتا اللتي وصلت سليمة من خلال تحديد افضل algorithm نقدر نحدد بيه الايرونر حصل فين بالضبط وازاي نعدله معنى reliable هو

How fast can data be reliably transmitted over a noisy channel? (For a noisy channel: The best possible error correction that can be achieved)

والاجابة على السؤال ده طلعت لنا نظرية اسمها ال Channel coding وتعريفها هو

Channel coding theorem: The **maximum** rate at which data can be **reliably transmitted** is the channel capacity of the channel.

■ Who started the information theory ?

Claude Shannon (1916 - 2000). The father of information theory.

Article in 1948: The mathematical theory of communication.

■ Applications : بتستخدم في ايه المادة؟

Data compression

Error correcting coding

Transmission and modulation

Image processing: texture

Information security ...



What is information?

A sequence of symbols that can be interpreted as a message.

يُعرف ال info على أنها تتابع من الرموز ينتج لنا رسالة في النهاية , وده تعريف مش مهم في مادتنا لان تعريف ال info في مادة ال information theory هو

It resolves uncertainty. Information is what you get when your uncertainty about something is reduced.

الكلام اللي جاي نظري كتير بس مهم علشان تفهم القانون في النهاية اتعمل ازاى

Uncertainty معناها الشك وهو ضد المعرفة او ال information , فلما تزيد معرفتك بحاجه يقل شكك فيها مثال : لما بترمي جنيه مش بتبقى عارف النتيجة هتبقى ملك ولا كتابة وهنا يكون ال uncertainty عالي وال info قليلة ولكن لما بترميه ويطلعك وليكن كتابة هنا بتزيد ال info و يقل ال uncertainty

Loss of uncertainty = Gain in information.

The more uncertain an event is, the more information is required to resolve uncertainty of that event.

Amount of information

ازاي احسب كمية المعلومات اللي انا اعرفها حاليا واللي بتطلع من مصدر معين وليكن الجنيه مثلا لنفترض ان الجنيه اسمه X والاحتمالات الممكنة لما ارمي الجنيه اسمها M وهنرمز لكمية المعلومات اللي هكتسبها لما اعمل التجربة بالرمز $L(X)$

$$I(X) = \log(M)$$

$$I(X) = \log_2 2 = 1 \text{ bit} \quad \text{فبقي ناتج المعادلة للمثال بتاعنا هو}$$

The information received from an information source equals the logarithm of the number of possible outcomes.

The unit used to measure $I(X)$ depends on the base of the logarithm.

وفي الكورس ده هيكون ال base بتاع ال log هو 2 دايما , لاننا بنتعامل مع binary numbers واللي هي احتماليين بس يا صفر يا واحد

طيب لو كان عندي مصدر بيطلع output واحد بس (مش ملك وكتابة زي مثال الجنيه) فلما نيجي نعوض في القانون هنلاقه كالاتي

$$I(X) = \log_2 1 = 0 \text{ bit}$$

و صفر هنا نتيجة بديهية لاني بحسب كمية المعلومات اللي هكتسبها زي ما شرحنا فوق وهنا انا كنت عارف اوريدي ال output قبل ما اجرب التجربة لان التجربة اساسا ملهاش غير output واحد فالباتلي كمية اللي اكتسبتها لما عملت التجربة كانت ب صفر

مثال ثاني بس المرادي على حجر النرد مش الجنيه

A dice has six possible outcomes.



$$I(X) = \log_2 6 = 2.585 \text{ bits}$$

Since the last example **has more outcomes**, the **amount of uncertainty increases**. Therefore, the amount of information is bigger than that of the coin examples.

we will use (log) instead of (log2), while we actually mean (log2).

مش log ال 10 زي ما احنا متعودين

$$\log_2 a = \frac{\log a}{\log 2} \approx 3.32 \times \log a$$

بسطنا القانون باستخدام الخاصية رقم 6 في جدول خواص اللوغاريتمات
3.32 جت منين ؟ جت من هنا

$$\frac{1}{\log 2} = 3.32$$

In all the previous examples, the outcomes are **equally likely**.

يعني احتمال ظهور الملك كان 0.5 واحتمال ظهور الكتابة برضو 0.5 , واحتمال ظهور اي رقم في النرد كان ب 1/6
طيب لو كانوا مش equally likely ؟ يعني احتمال ظهور الملك مثلا 0.99 والكتابة 0.01
وقتها هيبقى القانون كالاتي

$$I(a_i) = \log\left(\frac{1}{p_i}\right)$$

وهتبقى نواتج المثال اللي فوق بفرض ان الملك a_1 والكتابة a_2

$$I(a_1) = \log\left(\frac{1}{0.99}\right) = 0.014 \text{ bit}$$

$$I(a_2) = \log\left(\frac{1}{0.01}\right) = 6.644 \text{ bit}$$

The outcomes that are **less common** give us **more information** and vice versa.

وهنلاحظ ان احتمال حدوث العنصر لما بيكون قليل بيطلع ناتج اكبر زي ما حصل مع الكتابة والعكس صحيح

طيب افرض عايزين نحسب ال amount of information للتجربة كلها مش للملك بس او الكتابة بس
وقتها هيبقى هنحسب ال amount of information للملك و للكتابة وبعدين نجيب للرقمين ال **average** او

amount of information entropy مش ال **expected value** , وهنا هنقول اننا بنحسب ال

Entropy is **the average number** of bits required to represent an information source.

كنا بنحسب ال **expected value** باننا نمسك كل amount of information معنا نضربه في الاحتمال بتاعه وبعدين نجيب مجموع النواتج دي ونقسمهم على عددهم , فيبقى قانون ال entropy كالاتي وهنرمز له بالرمز $H(X)$:

$$H(X) = \sum_{i=1}^M p_i \log\left(\frac{1}{p_i}\right)$$

وباستخدام الخاصية رقم 5 من خواص اللوغاريتمات نقدر نبسط القانون للشكل ده (افكر ان واحد على اي حاجة هي نفس الحاجة اس سالب واحد)

$$= - \sum_{i=1}^M p_i \log(p_i)$$

هناخد 3 امثلة على القانون ده كلهم تعويض مباشر وبس كده ♥

1. Fair coin with probabilities of $\{0.5, 0.5\}$

$$H(X) = 0.5 \times \log\left(\frac{1}{0.5}\right) + 0.5 \times \log\left(\frac{1}{0.5}\right) = 1 \text{ bit.}$$

2. Biased coin with probabilities of $\{0.75, 0.25\}$

$$H(X) = 0.75 \times \log\left(\frac{1}{0.75}\right) + 0.25 \times \log\left(\frac{1}{0.25}\right) = 0.811 \text{ bits.}$$

3. Unbalanced dice with probabilities of $\{0.1, 0.1, 0.1, 0.5, 0.1, 0.1\}$

$$H(X) = 5 \times (0.1 \times \log\left(\frac{1}{0.1}\right)) + 0.5 \times \log\left(\frac{1}{0.5}\right) = 2.161 \text{ bit}$$

توضيح لآخر مثال : ضربت في خمسة في الجزء الاولاني من المعادلة لان ال 0.1 متكررة عندي 5 مرات , وكنت اقدر انفذ عملية الجمع 5 مرات بدل ما اضرب في 5 بس الضرب اسهل من اني اكتب نفس المعادلة 5 مرات

summary of summary :

- The central problem in information theory is **efficient and reliable** transmission of data from a source to a destination.
- 1. **Source coding theorem:** The maximum rate at which data can be compressed losslessly is the entropy rate of the source.
- 2. **Channel coding theorem:** The maximum rate at which data can be reliably transmitted is the channel capacity of the channel.
- **Information** is sequence of symbols that can be interpreted as a message.
- **Information** resolves uncertainty. Information is what you get when your uncertainty about something is reduced.
- Loss of uncertainty = Gain in information. (**inverse relationship**)
- **Entropy** is the average number of bits required to represent an information source.

➤ **Amount of information equation:**

$$I(X) = \log(M)$$

➤ **Entropy equation:**

$$H(X) = \sum_{i=1}^M p_i \log\left(\frac{1}{p_i}\right) = - \sum_{i=1}^M p_i \log(p_i)$$

(فَأَمَّا الزَّبَدُ فَيَذْهَبُ جُفَاءً وَأَمَّا مَا يَنْفَعُ النَّاسَ فَيَمْكُثُ فِي الْأَرْضِ)