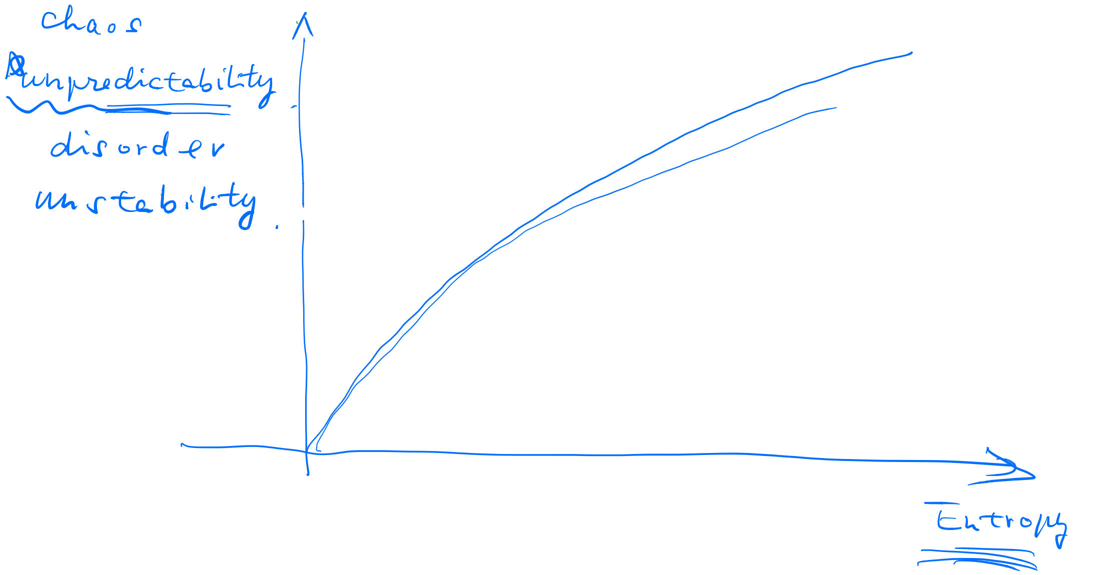
CPSC 4430/5440: Machine Learning

Lesson A05: Info Science

Outline:

1. Definition of Entropy.
2. 3 type of Entropy. — Shannon, Cross, KL.
3. Concepts of Shannon Entropy.
4. Decision Tree Based on Entropy.
   — info gain. (IG)
5. Cross Entropy.
6. Kullback_Leibler (KL) entropy.

Concept of Entropy.



Higher Entropy Value

⟶ More difficult for us to discover knowledge (extract)

What should we do:

⟹ Try to lower the Entropy of Data System

(

3 Types of Entropy.

① Shannon:

$$E(p) = -\sum_i p_i \cdot \log p_i$$

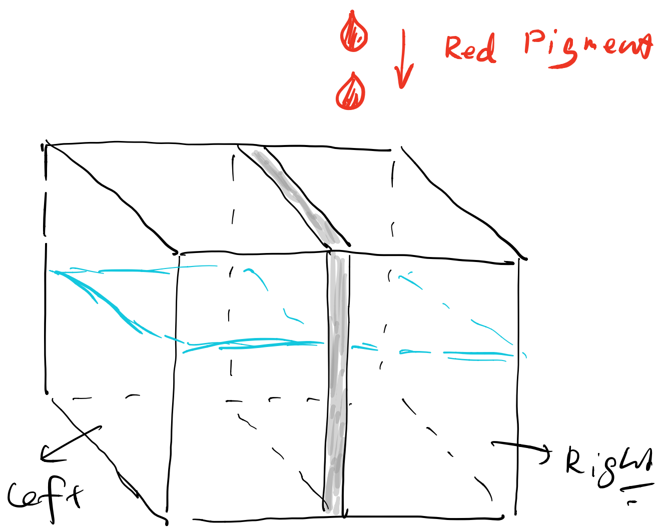
cross Entropy

$$H(p, q) = -\sum p_i \cdot \log_2 q_i$$

KL- entropy

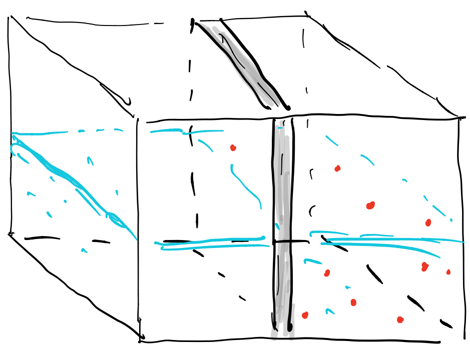
$$KL(p/q) = -\sum_i p_i \log_2 \left(\frac{p_i}{q_i}\right).$$

Shannon — Derivation from Thermodynamics.



A container is divided into 2 chambers.

A

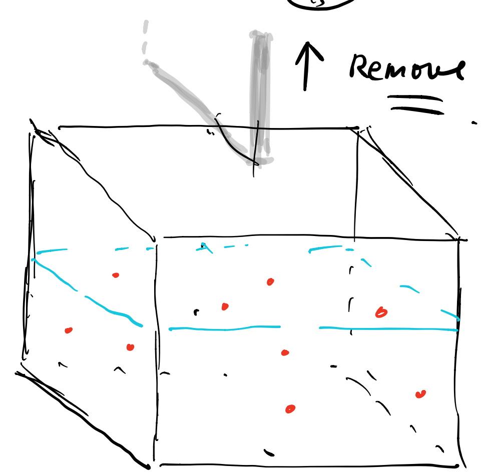


The probability of the pigment molecule appear on the Right Chamber.

$P = 100\%$.

in left $1-p = 0\%$

B



in Right: $P = 50\%$.

in Left: 50%.

$E(p)$

$= -p \cdot \log_2 p - (1-p)\log_2(1-p)$

$= -100\% * \log_2 1 - 0 * \log_2(0)$

$= 0.$

$E(p)$

$= -0.5 \cdot \log_2 0.5$

$\quad -0.5 \cdot \log_2 0.5$

$= 1.$

Lower Entropy Value

$\xrightarrow{\text{implies}}$ Higher Deterministic. (predictability)

Less stability.

Another Example to show Shannon Entropy.

Stock Investment

option	Probability of gain : loose	Shannon Entropy	(Deterministic) Difficulty scale to mak Decision
Stock A	100% : 0%	$-1 * \log 1 - 0 * \log 0 = 0$.	✓
B	75% : 25%	$-0.75 * \log_2 0.75 - 0.25 * \log_2 0.25 = 0.87$.	✓✗
C	25% : 75%	$-0.25 * \log_2 0.25 - 0.75 * \log_2 0.25 = 0.87$.	✓✗
D	0 : 100%	$-0 * \log 0 - 1 * \log 1 = 0$.	✓
E	50% : 50%.	$-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1$	✗

Larger Entropy.

→ Higher unpredictability.

→ ML aim to Decrease the Entropy of Data System

Construct Decision Tree based on Entropy.

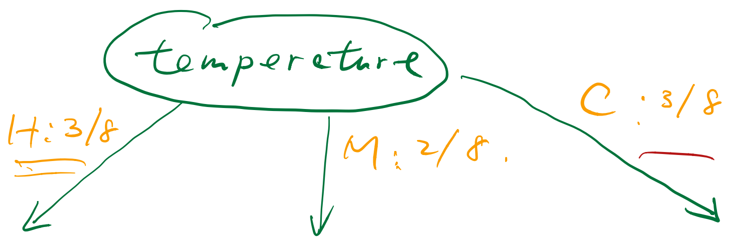
Outlook	Temperature	Humidity	Windy	play golf
Rainy	Hot	High	False	N
Rainy	Hot	H	True	N
Overcast	Hot	H	F	Y
Sunny	Mild	H	F	Y
Sunny	Cool	Normal	F	Y
Overcast	Cool	N	T	N
Rainy	Cool	N	T	Y
Rainy	Mild	H	F	N.

Shannon Entropy without classification

$$E(\text{play Golf}) = -\frac{1}{2}\cdot\log\frac{1}{2} - \frac{1}{2}\cdot\log\frac{1}{2}.$$

$$= 1.$$

Shannon Entropy based on temperature-oriented classification



Out	Humi	Wind	golf
			N
			N
			Y

O	H	W	golf
			Y
			N

O	H	w	golf
			Y
			N
			Y

$$E(\text{golf}, \text{temp} = H)$$

$$= -\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{2}{3} \cdot \log_2 \frac{2}{3}$$

$$= +0.8$$

$$E(\text{golf}, \text{temp} = M)$$

$$= -\frac{1}{2} \cdot \log \frac{1}{2} - \frac{1}{2} \cdot \log \frac{1}{2}$$

$$= 1$$

$$E(\text{golf}, \text{temp} = C)$$

$$= -\frac{1}{3} \cdot \log \frac{1}{3} - \frac{2}{3} \cdot \log \frac{2}{3}$$

$$= +0.8$$

$$E(\text{golf}, \text{temp})$$

$$= \frac{3}{8} * E(\text{golf}, \text{temp} = H) + \frac{2}{8} \cdot E(\text{golf}, \text{temp} = M) + \frac{3}{8} \cdot E(\text{golf}, \text{temp} = C)$$

$$= \frac{3}{8} * 0.8 + \frac{2}{8} * 1 + \frac{3}{8} * 0.8 = \underline{\underline{0.85}}$$

After categorizing Data
According temperature

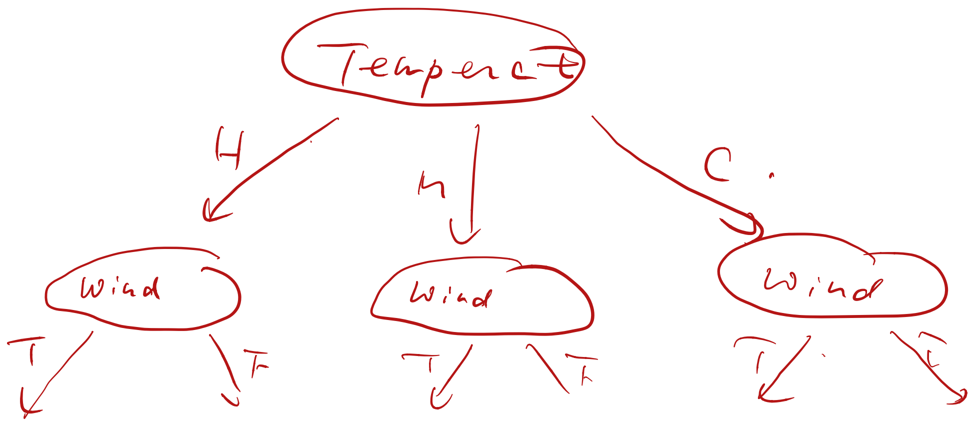
$E(\text{golf}, \text{temp}) = 0.85$

$E(\text{golf}) := 1.$

$\Downarrow$

We have Reduced the Entropy from $1 \rightarrow 0.85$

Information Gain (IG)
$= \underline{\underline{0.15}}.$



$E(\text{golf}, \underline{\underline{\text{temp}, \text{wind}}})$.

Cross Entropy

$$H(p, q) = -\sum_i p_i \log_2 q_i$$

Label	prediction
10%	15%
20%	23%
5%	10%
p	q

$$H(p, q) = -0.1 * \log_2 0.15 - 0.2 * \log_2 0.23 - 0.05 * \log_2 0.1$$

Kullback – Leibler (KL) D.

$$KL(p/q) = \sum_i p_i \log_2\left(\frac{p_i}{q_i}\right),$$

$$= H(p, q) - E(p),$$

Cross Entropy

Shannon Entropy