187 - Day 13

Wednesday, May 2, 2018 9:54 AM

Entropy = uncertainty of a random variable.
= average amount of information we gain by knowing the outcome of a R.V.
= average number of bits we need to store the outcomes of R.V

Day 13 – Properties of the Entropy

Elements of Information Theory

Definition 1. *The **entropy** of an event A is a measure of the uncertainty we feel about the occurrence of A.*

The entropy of a random variable X is given by

$$H(X) = \sum_{a} \mathbb{P}(X = a) \cdot \underbrace{\log_2 \left(\frac{1}{\mathbb{P}(X = a)} \right)}_{\text{amount of info we gain by knowing } X=a}$$

all outcomes

amount of info we gain by knowing $X=a$.

Definition 2. *The **entropy of two random variables** X and Y is*

$$H(X, Y) = \sum_{a,b} \mathbb{P}(X = a, Y = b) \cdot \log_2 \left(\frac{1}{\mathbb{P}(X = a, Y = b)} \right)$$

– the average amount of info we gain by learning the outcome of both X & Y

all outcomes $(X,Y) = (a,b)$

$\mathbb{P}(X=a \overset{\text{"and"}}{\cap} Y=b)$

Definition 3. *The **conditional entropy of the random variable** X given an **event** B is*

$$H(X|B) = \sum_{a} \mathbb{P}(X = a|B) \cdot \log_2 \left(\frac{1}{\mathbb{P}(X = a|B)} \right)$$

The additional amount of info we gain by learning X after we know that B has occurred.

Definition 4. *The **conditional entropy of the random variable** X given a **random variable** Y is*

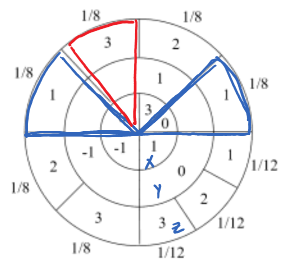
$$H(X|Y) = \sum_{b} \mathbb{P}(Y = b) \cdot H(X|Y = b)$$

sum over all the condition events

$\mathbb{P}$(condition event happens)

1

Example. Suppose that random variables X, Y, Z are obtained by spinning the wheel below, with X given by the innermost circle, Y given by the intermediate circle, and Z given by the outermost circle.



$$\mathbb{P}(Z=1 \cap X=0) = \frac{2}{8} \rightarrow \mathbb{P}(Z=1 \mid X=0) = \frac{\mathbb{P}(Z=1 \cap X=0)}{\mathbb{P}(X=0)} = \frac{2/8}{3/8} = \frac{2}{3}$$

$$\mathbb{P}(Z=2 \cap X=0) = 0 \rightarrow \mathbb{P}(Z=2 \mid X=0) = 0$$

↳ not on the wheel.

$$\mathbb{P}(Z=3 \cap X=0) = \frac{1}{8} \rightarrow \mathbb{P}(Z=3 \mid X=0) = \frac{\mathbb{P}(Z=3 \cap X=0)}{\mathbb{P}(X=0)} = \frac{1/8}{3/8} = \frac{1}{3}$$

last time

(a) Compute $H(X)$

(b) How many bits (of information) are required to store the results of 100,000 spins of Z? $= 100{,}000 \cdot H(Z)$

(c) Calculate the uncertainty of Z given that $X = 0$.

$$H(Z \mid X=0) = \sum_a \mathbb{P}(Z=a \mid X=0) \cdot \log_2\left(\frac{1}{\mathbb{P}(Z=a \mid X=0)}\right)$$

$$= \left(\frac{2}{3}\right) \cdot \log_2\left(\frac{1}{2/3}\right) + \left(\frac{1}{3}\right) \cdot \log_2\left(\frac{1}{1/3}\right)$$

$$\approx 0.9183$$

(d) Calculate $H(Z|Y) = \sum_y \mathbb{P}(Y=y) \cdot H(Z \mid Y=y)$

One way to do this: just compute all $H(Z|Y=-1)$, $H(Z|Y=0)$ and $H(Z|Y=1)$.

A faster way is to realize that Z and Y are independant random variables

Defn: 2 R.V. X and Y are independant if & only if.

$$\mathbb{P}(X=a \cap Y=b) = \mathbb{P}(X=a) \cdot \mathbb{P}(Y=b) \quad (*)$$

for all values $(X,Y) = (a,b)$.

You check that, under this wheel, Z & Y are independant - by checking $(*)$ holds for all pairs of values for (Z,Y)

Let X, Y be random variables. Then

$$H(Y|X) = H(Y) \Leftrightarrow X \text{ and } Y \text{ are independent.}$$

3

According to this, $H(Z|Y) = H(Z) = \log_2(3)$

(e) Calculate $H(X|Y,Z)$

Let $X, Y_1, Y_2, \ldots, Y_k$ be random variables then

$$H(X|Y_1, Y_2, \ldots, Y_k) = 0 \Leftrightarrow \underbrace{X = f(Y_1, \ldots, Y_k)}_{X \text{ is a fcn of } Y_1, Y_2, \ldots, Y_k}.$$

You don't gain any additional info for knowing X after knowing all Y_i's.

$H(X|Y,Z)$

Try to see if you can write X as a fcn of Y and Z.

If we have, for example: $X = Y^2 + 5Z^3$

then $H(X|Y,Z) = 0$

Suppose we learn the value of X and then we learn the value of Y. Then

$$H(X,Y) = H(X) + H(Y|X). = H(Y) + H(X|Y)$$

info about X, Y = info about X remaining info that we still don't know about Y, after knowing X

$$H(X_1, X_2, \dots, X_N) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_N|X_1, \dots, X_{N-1})$$

An observation: $\mathbb{P}(X=a \cap Y=b) = \mathbb{P}(X=a) \cdot \underbrace{\frac{\mathbb{P}(X=a \cap Y=b)}{\mathbb{P}(X=a)}}_{} = \mathbb{P}(X=a) \cdot \underbrace{\mathbb{P}(Y=b | X=a)}_{}$

Proof: $H(X,Y) = \sum_a \sum_b \mathbb{P}(X=a \cap Y=b) \cdot \log_2\left(\frac{1}{\mathbb{P}(X=a \cap Y=b)}\right)$

$$= \sum_a \sum_b \mathbb{P}(X=a \cap Y=b) \cdot \log_2\left(\frac{1}{\mathbb{P}(X=a)\cdot \mathbb{P}(Y=b|X=a)}\right)$$

$$= \sum_a \sum_b \mathbb{P}(X=a \cap Y=b) \cdot \log_2\left(\frac{1}{\mathbb{P}(X=a)}\right) + \sum_a \sum_b \mathbb{P}(X=a \cap Y=b) \cdot \log_2\left(\frac{1}{\mathbb{P}(Y=b|X=a)}\right)$$

$$= \sum_a \sum_b \mathbb{P}(X=a) \cdot \log_2\left(\frac{1}{\mathbb{P}(X=a)}\right) \cdot \mathbb{P}(Y=b|X=a) + \sum_a \sum_b \mathbb{P}(X=a) \cdot \mathbb{P}(Y=b|X=a) \cdot \log_2\left(\frac{1}{\mathbb{P}(Y=b|X=a)}\right)$$

$$= \underbrace{\sum_a \mathbb{P}(X=a) \cdot \log_2\left(\frac{1}{\mathbb{P}(X=a)}\right)}_{H(X)} \cdot \underbrace{\sum_b \mathbb{P}(Y=b|X=a)}_{1} + \underbrace{\sum_a \mathbb{P}(X=a) \cdot H(Y|X=a)}_{H(Y|X)}$$

5

Important inequalities

1. For a random variable X which takes only k values we always have

more uncertain

$$H(X) \leq \log_2(k)$$

with equality if and only if X takes all its values with equal probability

max uncertainty about X occurs when all its values are equally likely.

2. For any two random variables X and Y we always have

$$H(X|Y) \leq H(X)$$

and equality holds if and only if X and Y are independent

the amount of info we gain by learning X after knowing Y is less than the amount of info we would gain if we did not know Y.

3. For any two random variables X and Y we always have

$$H(X,Y) \leq H(X) + H(Y)$$

and equality holds if and only if X and Y are independent.

the amount of info. we gain by learning X & Y is less than the amount of info we gain if we learn these X, Y separately