Class no 3 :-

Signals → Carry Information

Comm → Reconstructing info at Rx end.

But what is information?

↳ Possible answers:

→ Anything interesting

→ Part of Data collected which is "useful"

→ What is interesting is something 'unexpected'

— Information of any outcome has to do with 'uncertainty' of that outcome.

'new' 'novel' 'unexpected' → Indicate 'high level of information.

Information deals with things/signals we don't know precisely but we are interested to find out

Say that X is a random quantity representing the signal of interest (at some pt of 'time' or 'space')

Example : $X \in \{0, 1\}$ & can take value 0 with probability p, & value 1 with prob $1-p$.

↳ General abstraction of many real world scenarios

X is also technically called a 'Random Variable'.

What is the information content in X ?

"Average uncertainty or surprise" in X ?

Information content in a particular event ∝ Inversely proportional to P(event)

(here there are 2 possible outcomes {0, 1} & 2 corresponding events $X=0$ & $X=1$)

For ex:

Information content in event "$X=0$" $\propto \frac{1}{P(X=0)}$

ditto "$X=1$" $\propto \frac{1}{P(X=1)}$

What fn can we assign to "Information content"?

→ Suppose X_1, X_2 are independent random variables

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) \quad \forall \text{ outcomes } x_1, x_2.$$

(x_1: some value; x_2: some value)

We expect the 'Joint information content in X_1 & X_2' = Sum of Individual Info content

→ Thus product of probabilities has to convert into sum of Information content

So for Info Content in event $X=x$, the $\log_2\left(\frac{1}{P(X=x)}\right)$ is a good choice of fn.

Base 2

Note: $\log\left(\frac{1}{P(X_1=x_1, X_2=x_2)}\right) = \log_2 \frac{1}{P(X_1=x_1)} + \log_2 \frac{1}{P(X_2=x_2)}$

if X_1, X_2 are independent

The 'Info content' in event $X = x$ ← same value which X can take

is assumed to be $\log \frac{1}{P(X=x)}$

So the 'average' info content ('expected' info content) 'mean'

Entropy of Random variable X denoted by $H(X)$ $\triangleq \sum_{x \in \text{all possible values taken by } X} p(X=x) \left[\log_2 \frac{1}{P(X=x)} \right]$

(Notation " $\triangleq$ " denotes that LHS is "defined as" RHS

'belongs to' (takes values in)

Lemma: Suppose $X_1 \in \mathcal{X}_1$ & $X_2 \in \mathcal{X}_2$

are independent random variables.

Then $H(X_1, X_2) = H(X_1) + H(X_2)$

where

$$H(X_1, X_2) \triangleq \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} P(X_1 = x_1, X_2 = x_2) \log \frac{1}{P(X_1 = x_1, X_2 = x_2)}$$

"Joint entropy of X_1 & X_2"

Class 4:

Proof: Use the fact that X_1, X_2 are indep & the fact that $\sum_{x_1 \in \mathcal{X}_1} P(X_1 = x_1) = 1 = \sum_{x_2 \in \mathcal{X}_2} P(X_2 = x_2)$

in defn of $H(X_1, X_2)$ and

prove the result. [Exercise for Monday, 31st May]

Proof :- Firstly, Independence of 2 RVs : Defn

ASIDE $\Leftarrow$ Two random variables X_1, X_2 are said to be independent if

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) \quad \forall\, x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2$$

$$H(X_1, X_2) = \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} P(X_1 = x_1, X_2 = x_2) \log_2 \frac{1}{P(X_1 = x_1, X_2 = x_2)}$$

ASIDE

Note: $P(X_1 = x_1, X_2 = x_2)$ = prob that $X_1 = x_1$ AND (simultaneously) $X_2 = x_2$

$$= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) \left[\log_2\left(\frac{1}{P(X_1 = x_1)}\right) + \log_2\left(\frac{1}{P(X_2 = x_2)}\right) \right]$$

because of independence of X_1 & X_2

$$= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) \log\left(\frac{1}{P(X_1 = x_1)}\right)$$

$$+ \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) \log \frac{1}{P(X_2 = x_2)}$$

$$= \sum_{x_1 \in \mathcal{X}_1} \log\left(\frac{1}{P(X_1 = x_1)}\right) \left[\sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) \right]$$

$$+ \sum_{x_2 \in \mathcal{X}_2} \log \frac{1}{P(X_2 = x_2)} \left[\sum_{x_1 \in \mathcal{X}_1} P(X_1 = x_1, X_2 = x_2) \right]$$

$\hookrightarrow$ use ASIDE C here from next page

$$= H(X_1) + H(X_2)$$

ASIDE (C)

Now

Claim: $\sum_{x_2 \in \mathcal{X}_2} P(\overbrace{X_1 = x_1, X_2 = x_2}^{A_{x_2}}) = P(X_1 = x_1)$

even if not independent

Whether X_1 is independent of X_2 or not

Proof of claim:

Suppose A & B are disjoint or mutually exclusive events
(for eg: $X_2 = x_2$, $X_2 = x_2'$ are m.e. if $x_2 \neq x_2'$)

$$P((X_2 = x_2) \cap (X_2 = x_2')) = 0.$$

$$P(A \cap B) = 0$$

$$P(A \cup B) = P(A) + P(B)$$

In general
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Note: $\left.\begin{array}{l} A = (X_1 = x_1) \cap (X_2 = x_2) \\ B = (X_1 = x_1) \cap (X_2 = x_2') \end{array}\right\}$ are disjoint (m.e) if $x_2 \neq x_2'$.

$\Rightarrow$ $P(A \cap B) = 0$ & $P(A \cup B) = P(A) + P(B)$

Now by same logic the events $(X_1 = x_1) \cap (X_2 = x_2)$ are disjoint for different values of $x_2 \in \mathcal{X}_2$.

Thus $\sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2)$

$$= \sum_{x_2 \in \mathcal{X}_2} P(\underbrace{\boxed{(X_1 = x_1) \cap (X_2 = x_2)}}_{A_{x_2}}) = \sum_{x_2 \in \mathcal{X}_2} P(A_{x_2})$$

$$= P\left(\bigcup_{x_2 \in \mathcal{X}_2} A_{x_2}\right) \quad \left[\text{as } A_{x_2} : x_2 \in \mathcal{X}_2 \text{ events are disjoint}\right.$$

$$\bigcup_{x_2 \in \mathcal{X}_2} A_{x_2} = \bigcup_{x_2 \in \mathcal{X}_2} \left[(X_1 = x_1) \cap (X_2 = x_2) \right]$$

$$= (X_1 = x_1) \cap \left[\bigcup_{x_2 \in \mathcal{X}_2} (X_2 = x_2) \right]$$

$$= (X_1 = x_1) \cap \left[\underbrace{X_2 \in \mathcal{X}_2}_{\text{true always!}} \right]$$

$$P\left(\bigcup_{x_2 \in \mathcal{X}_2} A_{x_2}\right) = P\Big(\underline{(X_1 = x_1)} \cap \underbrace{[X_2 \in \mathcal{X}_2]}_{\text{whole sample space itself}}\Big)$$

$$= P\Big((X_1 = x_1) \cap \underbrace{\Omega}_{\text{sample space}}\Big)$$

$$= P(X_1 = x_1)$$

Question:

Now what happens to $H(X_1, X_2)$ if X_1, X_2 are not independent?

For that we need conditional $\underline{\text{probability distributions}}$ $\to \{ P(X_1 = x_1), \text{ for } \forall x_1 \in \mathcal{X}_1 \}$

$$\boxed{P(X_2 = x_2 \mid X_1 = x_1) \triangleq \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)}}$$

$\mid$ → "given that"

"under the condition that"

"conditioned on the fact that"

Defn is valid only for $P(X_1 = x_1) \neq 0$.

Please check: $\sum_{x_2 \in \mathcal{X}_2} \boxed{P(X_2 = x_2 | X_1 = x_1)} = 1$

Check: $Q(X_2 = x_2) \triangleq P(X_2 = x_2 | X_1 = x_1)$

Q is a prob distribution on X_2 by itself

$\hookrightarrow \quad Q(X_2 = x_2) \geq 0 \quad \forall x_2 \in \mathcal{X}_2$

$$\sum_{x_2 \in \mathcal{X}_2} Q(X_2 = x_2) = 1$$

Defn:

$$H(X_2 | X_1) \triangleq \sum_{\substack{x_1 \in \mathcal{X}_1 \\ \text{such that } P(X_1 = x_1) \neq 0}} p(X_1 = x_1) H(X_2 | X_1 = x_1)$$

"Conditional Entropy of X_2 given X_1."

where

$$H(X_2 | X_1 = x_1) \triangleq \sum_{x_2 \in \mathcal{X}_2} P(X_2 = x_2 | X_1 = x_1) \log_2 \left(\frac{1}{P(X_2 = x_2 | X_1 = x_1)} \right)$$

Exercise: Show that

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

(Note $H(Y|X)$ may not in general $= H(X|Y)$)

Class no: 5

Before the proof, a bit of clarification regarding defns of entropy, cond entropy

$$H(X) = \sum_{\text{supp}(P_X)} p(X = x) \log \frac{1}{p(X = x)}$$

$\text{supp}(P_X) = \{ x \in \mathcal{X} : p(X = x) \neq 0 \}$

$P_X(x) \triangleq P(X = x)$

Sometimes we omit X and simply write $p(x)$

$\mathrm{supp}(f) \triangleq$ set of all elements in the domain of the fn at which fn value $\neq 0$

$\Rightarrow f: \mathcal{X} \to \mathcal{Y}$ then

$$\mathrm{supp}(f) = \{ x \in \mathcal{X} : f(x) \neq 0 \}.$$

$$\mathrm{supp}(P_X) = \{ x \in \mathcal{X} : P_X(x) \neq 0 \}$$

$$H(X) \triangleq -\sum_{x \in \mathrm{supp}(P_X)} P_X(x) \log_2 P_X(x)$$

$$= -\sum_{x \in \mathrm{supp}(p)} p(x) \log p(x)$$

$$H(X|Y) = \sum_{y \in \mathrm{supp}(P_Y)} P_Y(y)\, H(X|Y=y)$$

denotes $P(X=x \mid Y=y)$

where $H(X|Y=y) \triangleq -\sum_{x \in \mathrm{supp}(P_{X|Y})} P_{X|Y}(x|y) \log P_{X|Y}(x|y)$

Coming back to proof of exercise:

$H(X) + H(Y|X)$

$$= \sum_{x \in \mathrm{supp}(P_X)} p(x) \log \frac{1}{p(x)} + \sum_{x \in \mathrm{supp}(P_X)} p(x) \left(\sum_{y \in \mathrm{supp}(P_Y)} p(y|x) \log \frac{1}{p(y|x)} \right)$$

$\Rightarrow$ clarified below

$$= \sum_{x \in \mathrm{supp}(P_X)} \left(\sum_{y \in \mathrm{supp}(P_Y)} p(x,y) \right) \log \frac{1}{p(x)} + \sum_{x \in \mathrm{supp}(P_X)} \sum_{y \in \mathrm{supp}(P_Y)} p(x,y) \log \frac{1}{p(y|x)}$$

$$= \sum_{x \in supp(P_X)} \sum_{y \in supp(P_Y)} p(x,y) \left(\log \frac{1}{p(x)p(y|x)} \right)$$

$$= \sum_{x \ldots} \sum_{y \ldots} p(x,y) \log \frac{1}{p(x,y)} = H(X,Y)$$

$\rightarrow$ proof is complete

Clarification:

$$p(x) = \sum_{y \in \mathcal{Y}} p(x,y) = \sum_{y \in supp(P_Y)} p(x,y) + \sum_{y \in \mathcal{Y} \setminus supp(P_Y)} p(x,y)$$

For $y \in \mathcal{Y} \setminus supp(P_Y)$ what is $P_Y(y)$. It is $= 0$.

If $P_Y(y) = 0$ what abt $P_{X,Y}(x,y)$?

$$P_{X,Y}(x,y) = p(Y=y)\, p(X=x \mid Y=y)$$

$$= 0 \checkmark$$

$\Rightarrow$ This means green term above $= 0$

$$\Rightarrow \sum_{y \in \mathcal{Y}} p(x,y) = \sum_{y \in supp(P_Y)} p(x,y)$$

Some properties of entropy:-

Can we bound $H(X)$ from above & below

Claim: $0 \leq H(X) \leq \log_2 |\mathcal{X}|$

$\uparrow$ lower bound $\quad$ $\nwarrow$ upper bound

Proof for "$H(X) \geq 0$":

$$H(X) = \sum_{x \in supp(P_X)} p(x) \log_2 \frac{1}{p(x)}$$

$\log_2 \frac{1}{p(x)}$ ↳ always ≥ 0 since $p(x) > 0$ for any $x \in supp(P_X)$

$p(x) \geq 0$ since $p(x)$ is a prob

[$= 0$ only when $p(x) = 1$]

$\Rightarrow$ $H(X) \geq 0$. //

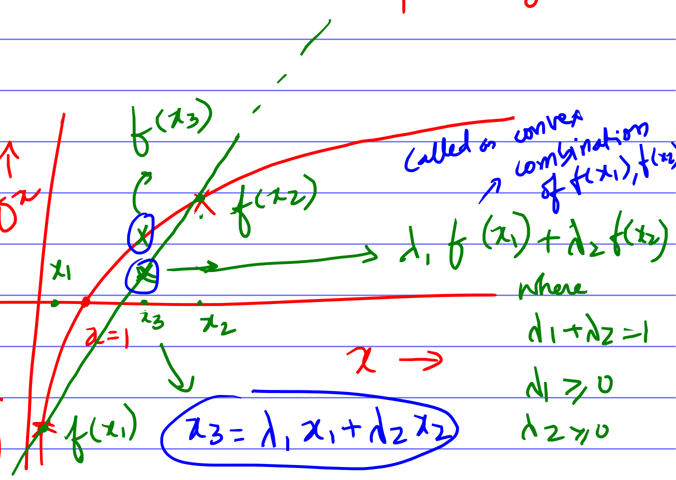
Q: When is $H(X) = 0$? $\rightarrow$ This happens if and only if

$X = x_0$ with $P_X(x_0) = 1$

& $P_X(x) = 0 \;\; \forall \; x \neq x_0$ $\Big\} \Leftrightarrow |supp(P_X)| = 1$.

Interpretation: Avg Uncertainty abt $X = 0$ as $X = x_0$ with probability 1.

Concave (Convex fns)



Jensen's inequality:-

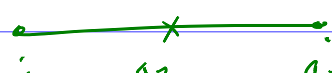
$f(\lambda_1 x_1 + \lambda_2 x_2)$
$\;\;\; \|$
$\rightarrow f(x_3) \geq \lambda_1 f(x_1) + \lambda_2 f(x_2)$

Concave fn

$\log(x_3) \geq \lambda_1 \log(x_1) + \lambda_2 \log(x_2)$

$\lambda_1, \lambda_2 \in [0,1]$

$\lambda_1 + \lambda_2 = 1$.



$a_3 = \lambda_1 a_1 + \lambda_2 a_2$

$\lambda_1 + \lambda_2 = 1$

$\lambda_1 \geq 0, \; \lambda_2 \geq 0$

For Convex $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$

$f(\sum \lambda_i x_i) \leq \sum \lambda_i f(x_i)$

Now we shall use the concavity property of log fn to obtain proof for $H(X) \le \log|\mathcal{X}|$.

Proof:

$$H(X) = \sum_{x \in supp(P_X)} p(x) \log \frac{1}{p(x)}$$

Let us assume that λ_x is another notation for $p(x)$

$$\boxed{\begin{array}{l} p(x) : x \in supp(p(x)) \\ \sum_{x \in supp(P_X)} p(x) = 1 \end{array}}$$

since this is true:

$$H(X) = \sum_{x \in supp(P_X)} \lambda_x \log\left(\frac{1}{p(x)}\right)$$

$\boxed{\text{Convex comb}}$ of $\log \frac{1}{p(x)} : x \in supp(P_X)$

$$\le \log\left(\sum_{x \in supp(P_X)} \lambda_x \frac{1}{p(x)}\right)$$

$$\le \log\left(\left|\underline{supp(P_X)}\right|\right) \le \log\underline{|\mathcal{X}|}$$

Question: When is $H(X) = \log|\mathcal{X}|$? (Use concavity & the ineq above to answer this)

Class 6:

When is Jensen's ineq satisfied with equality?

$\hookrightarrow$ Surely if f is a straight line Then this is true

So our worry is abt strictly concave (or strictly convex)

$\hookrightarrow$ eg: is $\log x$ curve (this curve is important for our discussion)

Claim:

Suppose $f(x)$ is strictly concave.

& $\lambda_1, \lambda_2 \neq 0$; $\lambda_1 + \lambda_2 = 1$

If: $f(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 f(x_1) + \lambda_2 f(x_2)$

then $\underline{x_1 = x_2}$.

More generally

If $\underline{\underline{\lambda_i \neq 0}}$, $\sum_{i=1}^{n} \lambda_i = 1$

& Jensen's holds with equality,

then $x_1 = x_2 = \ldots = x_n$.

Proof is easy using defn of strictly concave

Applying this to $H(X) \le \log|\mathcal{X}|$, looking at proof

$$H(X) = \sum_{x \in supp(P_X)} \lambda_x \log \frac{1}{p(x)} \le \log|supp(P_X)|$$

Suppose equality holds,

Then

$\Rightarrow$ by above Claim, we must have

$$\frac{1}{p(x)} = \text{const } c \ \forall x \in supp(P_X)$$

$\Rightarrow$ But we know $\sum_{x \in supp(P_X)} p(x) = 1$

$$\Rightarrow \text{const } c = p(x) = \frac{1}{|supp(P_X)|}$$

$H(X) = \log|supp(P_X)|$. If $|supp(P_X)| = |\mathcal{X}|$

Then $\Rightarrow \boxed{p(x) = 1/|\mathcal{X}|, \ \forall x \in \mathcal{X}}$.

This is called as a Uniform prob distribution

" We also then say the outcomes are equally likely

$\Rightarrow$ We have just proved that if $H(X) = \log_2 |\mathcal{X}|$,

$\rightarrow$ then this can be true only if P_X is uniform.

Thus it is easy to see

Lemma: $H(X) = \log_2 |\mathcal{X}|$ iff (if & only if) P_X is uniform.

["necessary and sufficient condition" for $H(X) = \log_2 |\mathcal{X}|$ is P_X = uniform]

Defn:

Relative Entropy (or) Information Divergence (or) Kullback-Leibler Divergence

Suppose there is a random X for which we have two probability distributions P_X & q_X.

Then the R-E or I-D or K-L divergence,

$$D(P_X \| q_X) \triangleq \sum_{x \in \mathrm{supp}(P_X)} p(x) \log \frac{p(x)}{q(x)}.$$

(Notice that $D(p\|q) \neq D(q\|p)$ in general)

[$D(p\|q)$ is a 'kind-of' a distance measure between distributions p, q]

Claim: $D(p\|q) \geq 0$.

Proof: $$D(p\|q) \doteq -\sum_{x \in supp(P_X)} p(x) \log \frac{q(x)}{p(x)}$$

$$\geq -\log \left(\sum_{x \in supp(P_X)} p(x) \frac{q(x)}{p(x)} \right)$$

$$\geq -\log \left(\sum_{x \in supp(P_X)} q(x) \right)$$

$$\geq 0 \quad \text{since } \sum q(x) \leq 1$$

Question: When is $D(p\|q) = 0$?

What we get by applying equality condition from Jensen's is

$$\Leftarrow \frac{p(x)}{q(x)} = \text{constant } c \text{ for } \forall x \in supp(P_X)$$

$$\Leftarrow 1 = \sum_{x \in supp(P_X)} p(x) = \sum_{x \in supp(P_X)} c\, q(x), \quad \forall x \in supp(P_X)$$

This together will mean that $c = 1 \Rightarrow p(x) = q(x)$ $\forall x \in supp(P_X)$

Thus by showing only if by above & if condition (straightforward),

$$\Rightarrow D(p\|q) = 0 \text{ iff } P_X = q_X.$$

Going back to conditional entropy

this is a valid prob distri for X (Check: $\sum_x p(x|y) = 1$)

$$H(X|Y) = \sum_{y \in supp(P_Y)} p_Y(y) \underbrace{\sum_{x \in supp(P_{X|Y})} p_{X|Y}(x|y) \log \frac{1}{p_{X|Y}(x|y)}}_{H(X|Y=y)}$$

(Conditional) Entropy in X given Y.

Claim: $0 \le H(X|Y) \le H(X)$

Proof: $H(X|Y) \ge 0$ is true because

$H(X|Y=y) \ge 0$ ← usual defn of entropy of X but with distb $P_{X|Y=y}$

& $p(y) \ge 0$.

$H(X) - H(X|Y)$

$$= \sum_{x \in supp(P_X)} p(x) \log \frac{1}{p(x)} - \sum_{\substack{x \in supp(P_X) \\ y \in supp(P_Y)}} p(x,y) \log \frac{1}{p(x|y)}$$

$$= \sum_{\substack{x \in supp(P_X) \\ y \in supp(P_Y)}} p(x,y) \log \frac{1}{p(x)} - \left(\quad = \quad \right)$$

because $\sum_{y \in supp(P_Y)} p(x,y) = p(x)$

$$= \sum_{\substack{x \in supp(P_X) \\ y \in supp(P_Y)}} p(x,y) \log \left(\frac{p(x|y)\, p(y)}{p(x)\, p(y)} \right)$$

$$= \sum_{\substack{x \in supp(P_X) \\ y \in supp(P_Y)}} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \rightarrow 1$$

Note that $\left\{ p(x,y) \,;\, \begin{matrix} x \in supp(P_X) \\ y \in supp(P_Y) \end{matrix} \right\}$.

is a valid joint distribution on X, Y

$\left(p(x,y) \geq 0 \,,\, \sum_{x \in supp(P_X)} \sum_{y \in supp(P_Y)} p(x,y) = 1 \right)$

What abt $\left\{ \underbrace{p_X(x)\, p_Y(y)}_{}^{q(x,y) \triangleq} : x \in supp(P_X), y \in supp(P_Y) \right\}$?

Check that this is also valid joint distribution of X, Y.

$\Rightarrow$ (1) becomes

$$H(X) - H(X|Y) = D\Big(p(x,y) \,\Big\|\, \overbrace{p_X(x)\, p_Y(y)}^{q(x,y)}\Big)$$

≥ 0 (already shown as $D(p\|q) \geq 0$)

$\Rightarrow H(X|Y) \leq H(X)$ // proof of claim is over

What is the significance of the above inequality?

If $\underline{H(X|Y) = 0}$ what does it mean? (in terms of P_X, P_Y)

If $\underline{H(X|Y) = H(X)}$, what does it mean?

$\hookrightarrow$ Equality in Jensens