Prelim: March 17 $7^{30} - 9^{30}$

→ Double sided A4 cheatsheet

→ Material up to today

HW3 posted → March 19

- Conditional MI

Recap: $I(X;Y) := D_{KL}(P_{XY} \| P_X \otimes P_Y)$

relation b/t MI and entropy

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$
$$= H(X) - H(X|Y)$$
$$= H(Y) - H(Y|X)$$

Let $P_{XYZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ and consider $P_{XY|Z}(\cdot|z)$, $z \in \mathcal{Z}$, then $P_{XY|Z}(\cdot|x) \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

We define

$$I(X;Y|Z=z) := D_{KL}\left(P_{XY|Z}(\cdot|z) \,\|\, P_{X|Z}(\cdot|z) \otimes P_{Y|Z}(\cdot|z)\right)$$

Definition (Mutual Information):

For $X, Y, Z \sim P_{XYZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$, the conditional MI between X and Y given Z is

$$I(X;Y|Z) = D_{KL}\left(P_{XY|Z} \,\|\, P_{X|Z} \otimes P_{Y|Z} \,|\, P_Z\right)$$

$$= \mathbb{E}_{z \sim P_Z}\left[I(X;Y|Z=z)\right]$$

Comments:

① $I(X;Y|Z)$ is a functional of P_{XYZ} (not of e.g. $P_{XY|Z}$)

② It can be shown that

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(XY|Z)$$

$$= H(X|Z) - H(X|YZ)$$

$$= H(Y|Z) - H(Y|XZ)$$

Definition (Markov Chain)

Let $X, Y, Z \sim P_{XYZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$. We say that $X \to Y \to Z$ forms a Markov chain if

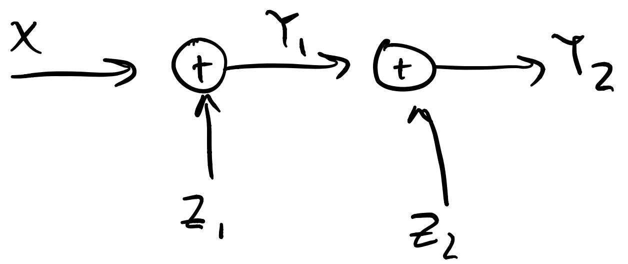
$$P_{XYZ} = P_X P_{Y|X} P_{Z|XY} = P_X P_{Y|X} \underbrace{P_{Z|Y}}_{\text{independent of past}}$$

Example

$$\left.\begin{aligned} X &\sim \mathcal{N}(0,\sigma^2) \\ Z_1 &\sim \mathcal{N}(0,\eta_1^2) \\ Z_2 &\sim \mathcal{N}(0,\eta_2^2) \end{aligned}\right\}\ \text{mutually independent}$$

Define $Y_1 = X + Z_1$

$$Y_2 = X + Z_1 + Z_2 = Y_1 + Z_2$$



$Y_2 \mid X, Z_1, Z_2$

$Y_2 \mid Y_1, Z_2$

Proposition (Equivalent Conditions)

The following are equivalent

$$\begin{aligned} X \to Y \to Z \iff\ & P_{XYZ} = P_X P_{Y|X} P_{Z|Y} \\ \iff\ & P_{XZ|Y} = P_{X|Y} P_{Z|Y} \\ \iff\ & P_{Z|XY} = P_{Z|Y} \\ \iff\ & X \perp\!\!\!\perp Z \mid Y \quad (X, Z \text{ conditionally independent } | Y) \\ \iff\ & Z \to Y \to X \\ & \cdots \end{aligned}$$

Proposition (More properties of MI)

Let $(X, Y, Z) \sim P_{XYZ}$. Then

① $I(X; Y | Z) \geq 0$ w/ equality iff $X \to Z \to Y$

② Chain Rule

(i) <u>Small</u> $I(X, Y; Z) = I(X; Z) + I(Y; Z | X)$

$$= I(Y; Z) + I(X; Z | Y)$$

(ii) <u>Full</u>: $I(X_1, \dots, X_n ; Z) = I(X_1 ; Y) + \sum_{i=2}^{n} I(X_i ; Y \mid X_1, \dots, X_{i-1})$

③ Data Processing Inequality

If $X \to Y \to Z$, then

$$I(X; Y) \geq I(X; Z)$$

w/ equality iff $X \to Z \to Y$.

④ If f is a bijection then

$$I(X; Y) = I(X; f(Y))$$

⑤ Concavity / Convexity

For $(X,Y) \sim P_{XY}$ denote $I(X;Y)$ as $I(P_X ; P_{Y|X})$.

Then

(i) For fixed $P_{Y|X}$, $P_X \mapsto I(P_X ; P_{Y|X})$ is concave

(ii) For fixed P_X, $P_{Y|X} \mapsto I(P_X ; P_{Y|X})$ is convex

Proof of proposition

① By def → complete details

② "Trivial"

$X \to Y \to Z$ MC

③ $I(X;Y,Z) = I(X;Y) + I(X;Z|Y) = I(X;Y) + 0$

$$= I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0}$$

w/ equality condition pops out!

$\Rightarrow I(X;Z) \leq I(X;Y)$

④ For any f deterministic

$$X \to Y \to f(Y)$$

$\xRightarrow{DPI} I(X;Y) \geq I(X;f(Y))$

But when f a bijection $X \to f(Y) \to Y$

$I(X;Y) \leq I(X;f(Y))$

$\Rightarrow I(X;Y) = I(X;f(Y))$

$\boxed{IV}$ Letter Typical Sequences

Overview for Binary Alphabets

- Let $\mathcal{X} = \{0, 1\}$ and consider its n-fold extension $\mathcal{X}^n$. This is the set of all binary sequences of length n. Elements of $\mathcal{X}^n$ are denoted as

$$x^n := (x_1, \dots, x_n) \in \mathcal{X}^n$$

Q1: How many sequences are in $\mathcal{X}^n$?

A1: $|\mathcal{X}^n| = |\mathcal{X}|^n = 2^n$

non-inclusive otherwise not interesting

Now let $P \in \mathcal{P}(\mathcal{X})$, i.e. $P = \mathrm{Ber}(\alpha)$, $\alpha \in (0, 1)$

Let $X \sim P$ and $X_1, X_2, X_3, \dots$ be iid copies of X. In other words: For $n \in \mathbb{N}$, $(X_1, \dots, X_n) \sim P^{\otimes n}$

Any time we draw $X^n := (X_1, \dots, X_n)$ we get a value in $\mathcal{X}^n$.

<u>Observation</u> For any $x^n \in \mathcal{X}^n$, we have $P^{\otimes n}(x^n) > 0$. Indeed if x^n has $k \le n$ 1's, then

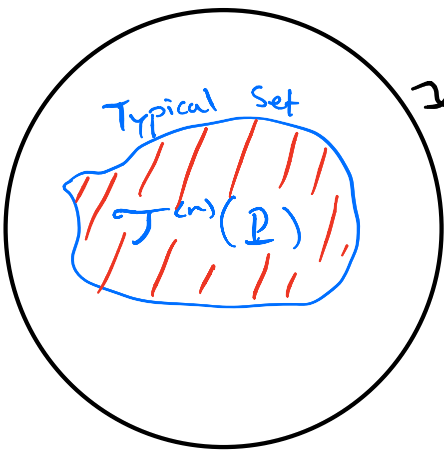
$$P^{\otimes n} = \alpha^k (1-\alpha)^{n-k} > 0 \quad \text{when} \quad \alpha \in (0,1)$$

Q2: Despite the above, what is the likely portion of 1's that we'll get by drawing X^n?

A2: Expect to see roughly $n\alpha$ 1's, $n(1-\alpha)$ 0's.

<u>Goal</u>: By exploiting the above observation, we will define for a subset of $\mathcal{X}^n$ ($\mathcal{T}^{(n)}(P)$), such that

(i) $|\mathcal{T}^{(n)}(P)| \ll |\mathcal{X}^n| \Rightarrow \dfrac{|\mathcal{T}^{(n)}(P)|}{|\mathcal{X}^n|} \xrightarrow[n\to\infty]{} 0$



$\mathcal{X}^n$ of size 2^n

Prob of $x^n \in \mathcal{T}^{(n)}(P)$ is almost 1

(ii) Set "absorbs almost all the probability" in the sense that

$$\lim_{n\to\infty} P^{\otimes n}\left(\mathcal{T}^{(n)}(P)\right) = 1$$

$$P^{\otimes n}\left(\{x^n \in \mathcal{X}^n \mid x^n \in \mathcal{T}^{(n)}(P)\}\right)$$

To formally set this up we need a few definitions.

Definition (Empirical Frequency):

Let $\mathcal{X}$ be discrete and $x^n \in \mathcal{X}^n$. Define

deterministic sequence

$$N_{x^n}(a) := \sum_{i=1}^{n} \mathbb{1}_{\{x_i = a\}} \qquad a \in \mathcal{X}$$

and then set $\nu_{x^n}(a) := \frac{1}{n} N_{x^n}(a)$, $a \in \mathcal{X}$, and observe that $\nu_{x^n}(a)$ is a valid pmf on $\mathcal{X}$. The pmf $\nu_{x^n}(a)$ is called the empirical frequency of $x^n \in \mathcal{X}^n$.