$\boxed{1}$ Probability

● Review

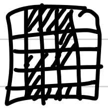
definition
$$
\begin{cases}
\Omega : \text{sample space} \\
E : \text{event} \\
\mathcal{F} : \text{set of events} \\
P : \mathcal{F} \to [0,1] = \{x \in \mathbb{R} : 0 \le x \le 1\}
\end{cases}
$$

Ex. Fair coin flip → probability of head, tail

$\Omega : \{H, T\}$ | $\mathcal{F} : \{\emptyset, \{H\}, \{T\}, \{H,T\}\}$

$Pr(\emptyset)=0 \quad Pr(\{H\})=\frac{1}{2} \quad Pr(\{T\})=\frac{1}{2} \quad Pr(\{H,T\})=1$

Ex 5x5 binary


$$= \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$\Omega = \{0,1\}^{5\times5}$

$|\Omega| = 2^{25}$

$|\mathcal{F}| = 2^{2^{25}}$

● Random variable

def $X : \Omega \to \mathbb{R}$

example: toss coin 5 times

$\Omega : \{H,T\}^5 = \{HHHHH, HHHHT, \cdots\}$

$X = \#$ of heads

$X(HHTHT) = 3$

$\boxed{2}$ Information Theory

($\approx$ Information Measures)

- **Surprisal** of x

def $S(x) = \log_2 \frac{1}{P_X(x)} = -\log_2 P_X(x)$

ex> $x = 1$ if earthquake
$\quad\;\; 0$ else

- def **Entropy** of random variable X

$$H(X) = \mathbb{E}[S(X)]$$

$$= \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] = \sum_{x \in \mathcal{X}} \left(\log_2 \frac{1}{P_X(x)}\right) \cdot P_X(x)$$

ex) X: binary random variable

Bern(p)

$$\Pr(X=1) = p$$

$$P_X(1) = p, \quad P_X(0) = 1-p$$

$$H(X) = \mathbb{E}\left[\log_2 \frac{1}{P_X(x)}\right] = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{1}{P_X(x)}$$

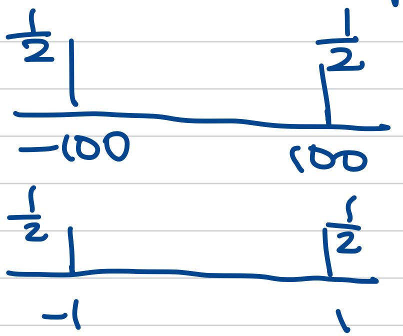
$$= p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$

Entropy = measure of uncertainty (randomness)

= amount of information

$\Rightarrow$ Units of Entropy : bits

std vs entropy



] same entropy
different standard deviation

* Entropy can also be used as image hardness

ex>
$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$$

A: p of dog
B: p of cat
C: p of fish
D: p of table

image1: $\begin{bmatrix} 0.8 \\ 0.05 \\ 0.10 \\ 0.05 \end{bmatrix}$ image2: $\begin{bmatrix} 0.6 \\ 0.2 \\ 0.1 \\ 0.1 \end{bmatrix}$

① $H(X)$ property : non-negativity

$H(X) \geq 0$

& $H(X) = 0 \Leftrightarrow X$: deterministic

$(P_r(X = x) = 1)$

Ex> $p_X(x) = \left(\frac{1}{2}\right)^x$ ($x \geq 1$, integer)

Q. $H(X) = ?$

A. $H(X) = \mathbb{E}\left[\log_2 \frac{1}{p_X(x)}\right] = \mathbb{E}[\log_2 2^x] = \mathbb{E}[x] = \sum_{x=1}^{\infty} x \cdot \left(\frac{1}{2}\right)^x$

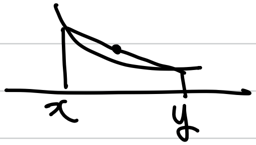
$$S(p) = \sum_{x=1}^{\infty} p^x = \frac{p}{1-p} \quad \text{when } |p|<1$$

$$\frac{dS}{dp} = \sum_{x=1}^{\infty} x \cdot p^{x-1} = \frac{d}{dp}\left(-1 + \frac{1}{1-p}\right) = \frac{1}{(p-1)^2} = 4$$

$$\therefore \sum_{x=1}^{\infty} x \cdot p^x = 4 \cdot p = 2$$

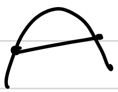
- def. **Convexity (Concavity)**

$f: \mathcal{X} \to \mathbb{R}$ is convex if $\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$ $\forall\, 0 \leq \lambda \leq 1,\ \forall x, y \in \mathcal{X}$



: convex

$f: \mathcal{X} \to \mathbb{R}$ is concave if $\lambda f(x) + (1-\lambda) f(y) \leq f(\lambda x + (1-\lambda) y)$



: concave

ex) $X \sim \text{Bern}(p)$

$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = H_2(p)$: concave.

- **Jensen's Inequality**

$f: \mathcal{X} \to \mathbb{R}$ concave

$\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$

$\Rightarrow$ max Entropy when uniform

- **Mismatch**

$H(X) = \mathbb{E}[\log \frac{1}{P_X(x)}] = \sum_x P_X(x) \frac{1}{\log \frac{1}{P_X(x)}}$

Suppose! $X \sim Q$

$H(X) \leq \mathbb{E}_{P(X)}[\log \frac{1}{Q(X)}]$

$\Longrightarrow$ **KL-Divergence**

• KL- Divergence (= relative entropy) → non negative by mismatch

<u>def</u>

$$D(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$$

* Not distance or metric but divergence

i) unstable : certain event ⇒ divergence is ∞

ii) symmetry X

*Earth Mover dist (= Wasserstein dist) : symmetric 있음

—— generate realistic image

GAN : true distribution img $= P_X(r.v)$

→ image Q

minimize dist between P_X and Q

• Cross Entropy

<u>def</u> $\sum_{x\in\mathcal{X}} \underbrace{P_X(x)}_{\text{true}} \log \underbrace{\frac{1}{Q(x)}}_{\text{예측값}}$

<u>indep def</u>) X & Y are indep

iff $P_{X,Y}(x,y) = P_X(x)\,P_Y(y)\ \forall x,y$

<u>def</u> **Random Vector**

$X^n = (X_1, X_2, \cdots, X_n)$ $x^n = (x_1, \cdots, x_n)$

joint pmf

$P_{X^n}(x^n) = \Pr(X^n = x^n) = \Pr(X_1 = x_1, \cdots X_n = x_n)$

Diffusion Overview

input $\rightarrow$ G $\rightarrow$ image $\quad$ true

$Q \approx P_X$

randomness

diffusion

$\boxed{\text{train process}}$

input $\rightarrow$ G $\rightarrow Q(x^n | c)$

Prompt $\rightarrow$ c

MNIST $Q(x^n)$

MNIST 7 $Q(x^n | 7)$

• Joint Entropy

def $H(X_1, X_2) = \sum_{(x_1, x_2)} P_{X_1, X_2}(x_1, x_2) \log \frac{1}{P_{X_1, X_2}(x_1, x_2)}$

$X_1 \in \mathcal{X}_1 = \{1, 2, \cdots, M_1\}$

$X_2 \in \mathcal{X}_2 = \{1, 2, \cdots, M_2\}$

$\Rightarrow$ total uncertainty of pair (X_1, X_2)

: total amount of information

Special case : $X_1 \perp\!\!\!\perp X_2 \Leftrightarrow P_{X_1 X_2}(x_1, x_2) = P_{X_1}(x_1) P_{X_2}(x_2)$

① $H(X_1, X_2) = H(X_1) + H(X_2)$ (when indep)

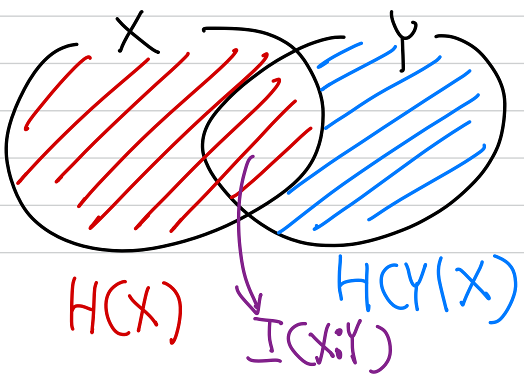
② $H(X_1, X_2) = H(X_1) + \underline{H(X_2 | X_1)}$

= Conditional Entropy

• Conditional Entropy

def $H(X|Y)$ = amount of info in X when Y is given?

$= \mathbb{E}[\log \frac{1}{P_{X|Y}(X|Y)}]$



I(X:Y) : Mutual Information

↳ nonnegative

↳ symmetric

• Data processing inequality

① $f : \mathcal{X} \to \mathbb{R}$

$H(X) \geq H(f(X))$

② $f : \mathcal{X} \to \mathbb{R}$

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

③ Markovity : seq of r.v

Markov triplet $X - Y - Z$: iff X & Z are indep given Y

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

• Random Process

① iid

② Markov Process (1st order)

$\underline{\text{def}}$ $P_{X_i | X_{i-1} \cdots X_1} = P_{X_i | X_{i-1}}$

(namely, Probability in ith stage only depend on previous stage ((i-1)th process)

③ kth order Markov Process

def $P_{X_i | X_{i-1} \cdots X_1} = P_{X_i | X_{i-1} \cdots X_{i-k}}$

④ **Stationary**

$X_1^n = X_{i+1}^{i+n} \quad \forall i, n$

- ## Continuous random variable

$F_X(x) = \Pr(X \le x)$: cumulative distribution function

$f_X(x) = \dfrac{dF_X(x)}{dx}$: probability density function

Jacobian?

If $Y = h(X) \Rightarrow |dy\, f_Y(y)| = |dx\, f_X(x)| =$ probability

$$\rightarrow f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|\frac{\partial h^{-1}(y)}{\partial y}\right| f_X(h^{-1}(y))$$

$(Y_1, Y_2) = H(X_1, X_2)$

$$f_{Y_1,Y_2}(y_1, y_2) = \underbrace{\left|\frac{dx_1 dx_2}{dy_1 dy_2}\right|}_{:= \det\left(\frac{\partial H^{-1}}{\partial y}\right)} f_{X_1,X_2}(x_1, x_2)$$

$:= \det\left(\dfrac{\partial H^{-1}}{\partial y}\right)$: Jacobian

- Gaussian Distribution

$$\underline{\text{def}}\quad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Jointly Gaussian $\sim N(\mu, \Sigma)$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

$$f_{X_1, X_2} = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}(X-\mu)\Sigma^{-1}(X-\mu)\right)$$

$$f_{X^n}(x^n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x^n-\mu)^T\Sigma^{-1}(x^n-\mu)\right)$$

• Differential Entropy

when discrete: $D(f \| g) = \mathbb{E}_f\left[\log \frac{f(X)}{g(X)}\right] = \int f(x) \log \frac{f(x)}{g(x)} dx$; ratio

when continuous: $D(P \| Q) = \mathbb{E}_P\left[\log \frac{P(X)}{Q(X)}\right] = \sum_x p(x) \log \frac{P(x)}{Q(x)}$; quantity

→ In continuous case

differential entropy $h(X) := \mathbb{E}_{f_X}\left[\log \frac{1}{f_X(X)}\right]$

Properties

① $h(X) \neq h(aX) = h(X) + \log|a|$

② $h(X_1, X_2) = h(X_1) + h(X_2 | X_1)$

③ $I(X;Y) = h(X) + h(Y) - h(X,Y)$

$h(X,Y) = \mathbb{E}_{f_{X,Y}}\left[\log \frac{1}{f_{X,Y}(X,Y)}\right]$