@Entropy

=> 정보의 단위

$$H(x) = -p(x)\log(p(x))$$

@KL divergence

=> 어떤 확률분포 p(x), p(y)가 있을 때, 이 둘의 차이 정의

$$KL(p|||q) = -p(x)\log[\frac{q(x)}{p(x)}]$$

@Cross-Entropy

=> 딥러닝에서 오차는 Cross_entropy로 계산하고 이것을 줄여나간다.

=> Cross_entropy가 최소가 되려면 X == Y인 경우이며, 즉 분포 X와 Y를 동일하게 맞춰나간다.

$$\begin{aligned} H(X,Y) &= H(X) + KL(X|||Y) \\ &= -p(x)\log(p(x)) - p(x)\log(\frac{q(x)}{p(x)}) \\ &= -p(x)\log(p(x)) - p(x)\log(q(x)) + p(x)\log(p(x)) \\ &= -p(x)\log(q(x)) \end{aligned}$$

@Mutual Information(두 변수 X, Y 사이의 상호 의존성 측정)

=> 다른 랜덤 변수(Y)를 통해 하나의 랜덤 변수(X)에 대해 얻은 정보량.

=> 만약 X, Y가 서로 독립이라면 정보량 == 0

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

@Conditional Entropy

=> H(Y|X=x) X 분포를 따르는 x의 Condition 일 때, Y의 정보량.

=> over all possible values x that X may take. 즉 SUM을 해줘야 함.

=> 아래 식은 이해는 못함..

$$\begin{aligned} H(Y|X) &= \sum_{x \in X} p(x) H(Y|X=x) \\ &= -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x)\log(p(y|x)) \end{aligned}$$

@Generative Adversarial Network Object Function

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_{data}}[\log D(x)] + E_{z\sim noise}[\log(1-D(G(z)))]$$

@InfoGAN

=> G 입장에서는 아래 식을 minimize 해야 하므로, 뒤에 붙은 Term을 Maximize 하면 됨.

=> c와 G(z,c) 사이의 상호 정보량은 서로 독립이면 0이다. 따라서 c와 G(z,c)를 상호 의존적으로 학습.

=> 이렇게 안하고 단순히 G를 G(z) 대신 G(z,c)만 사용하면, noise를 확장한 것과 의미상 같음.

$$\min_{G}\max_{D} V_I(D,G) = V(D,G) - \lambda I(c;G(z,c))$$

Conditional Entropy 공식 활용

$$I(c;G(z,c)) = H(c) - H(c|G(z,c)) \qquad (1)$$

$$= H(c) + E_{x\sim G(z,c)}[\boxed{E_{c'\sim P(c|x)}[\log(P(c'|x))]}] \qquad (2)$$

$$= H(c) + E_{x\sim G(z,c)}[\boxed{KL(P(c'|x)\|\|Q(c'|x)) + E_{c'\sim P(c|x)}[\log(Q(c'|x))]}] \qquad (3)$$

$$= H(c) + E_{x\sim G(z,c)}[\underbrace{KL(P(c'|x)\|\|Q(c'|x))}_{\geqq 0} + E_{c'\sim P(c|x)}[\log(Q(c'|x))]] \qquad (3)$$

$$\geqq H(c) + E_{x\sim G(z,c)}[E_{c'\sim P(c|x)}[\log(Q(c'|x))]] \qquad (4)$$

$$\geqq H(c) + E_{c\sim P(c), x\sim G(z,c)}[\log(Q(c|x)] \qquad (5)$$

$$= L_I(G,Q)$$

(2) -> (3) 유도 과정

$$KL(P(c'|x)\|\|Q(c'|x)) = -E_{c'\sim P(c|x)}\log\left(\frac{Q(c'|x)}{P(c'|x)}\right)$$

$$= -E_{c'\sim P(c|x)}\log(Q(c'|x)) + E_{c'\sim P(c|x)}\log P(c'|x)$$

$$\therefore \boxed{KL(P(c'|x)\|\|Q(c'|x)) + E_{c'\sim P(c|x)}[\log(Q(c'|x))]} == \boxed{E_{c'\sim P(c|x)}[\log(Q(c'|x)]}$$

(4) -> (5) 유도 과정

$P(c|x)$를 제거 했지만, 기댓값인 $E_{c'\sim P(c|x)}$ 에는 아직 $P(c|x)$가 남아있으므로, 이것을 제거해야함

$LEMMA: E_{x\sim X,\, y\sim Y|x}[f(x,y)] == E_{x\sim X,\, y\sim Y|x,\, x'\sim X|y}[f(x',y)]$ 를 따름.

$$\therefore \min_{G}\max_{D} V_{InfoGAN}(D,G,Q) = V(D,G) - \lambda L_I(G,Q)$$