[KL Divergence]

- Cross entropy : 예측과 달라서 생기는 정보량

ex) Binary cross entropy $= -\{y\log\hat{y} + (1-y)\log(1-\hat{y})\}$

$\hookrightarrow \ominus \sum_{y\in\{0,1\}} -T(y)\log P(\hat{y})$

$T(y)$: target $(y\in\{0,1\})$

$P(y)$: predicted

- if $y=1, \hat{y}=1$:
  $BCE = 0$
- if $y=1, \hat{y}=0$:
  $BCE = -\{-\infty+0\} = \infty$
- if $y=0, \hat{y}=1$:
  $BCE = -\{0+-\infty\} = \infty$
- if $y=0, \hat{y}=0$
  $BCE = 0$

$\downarrow$ multiclass (n)

Cross entropy $\sum_{y=1}^{n} -T(y)\log P(\hat{y})$

$\{y\in\{1,\cdots,n\}$

정상적인 주사위 (Normal)

X	1	2	3	4	5	6
$P(x=X)$	1/6	1/6	1/6	1/6	1/6	1/6

비정상적인 주사위 (Abnormal)

X	1	2	3	4	5	6
$Q(x=X)$	1/2	1/10	1/10	1/10	1/10	1/10

Cross entropy for Abnomal dice

$$: -\sum_y T(y)\log P(\hat{y}) = -\{1\times\tfrac{1}{6}\log\tfrac{1}{2} + 5\times\tfrac{5}{6}\log\tfrac{1}{10}\} = -\log\left(\tfrac{1}{2}\right)^{\frac{1}{2}}\times\left(\tfrac{1}{10}\right)^{\frac{25}{6}}$$

Cross entropy for Normal dice

$$: -\sum_y T(y)\log P(\hat{y}) = -6\times\tfrac{1}{6}\log\tfrac{1}{6} = -\log\tfrac{1}{6}$$

Cross entropy for Abnormal dice > Cross entropy for Normal dice

KL Divergence = Kullback - Leibler Divergence

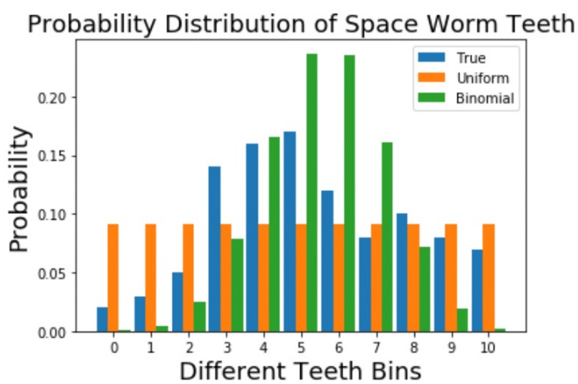
(= Relative Entropy, Information Gain)

↳ A way of measuring the matching between two distributions

정보량 (entropy) : 불확실성의 측정 (H)

p_i : 사건 발생 확률

$$H = \sum_i p_i \log_2 \left(\frac{1}{p_i}\right)$$



↳ How do we quantitatively decide which one is the best?

A way to measure the matching between each approximated distribution and the true distribution.

⇒ KL divergence

$$D_{KL}(p\|q) = \sum_{i=1}^{N} p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right)$$

$q(x)$: approximation

$p(x)$: true distribution we're interested in matching $q(x)$ to.

$0 \le D_{KL}(p\|q) \le \infty$

Lower the KL divergence value, the better we have matched the true distribution with our approximation.