Inclass 20: Information Theory

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

좋은값

Negative logarithm of the likelihood, which gives the cross-entropy error function

두개의 확률분포: 차이도

(t_n과 y_n

min

$$E(\mathbf{w}) = -\log p(\mathbf{t} \mid \mathbf{w}) = -\sum_{n=1}^{N} \{t_n \log y_n + (1 - t_n)\log(1 - y_n)\} \quad (1)$$

y_n이 t_n에 가까울수록 $E(w)$는 작아짐

y_n을 t_n처럼 만들고 싶었자나

Taking the gradient of the error function, we obtain

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (y_n - t_n)\mathbf{x}_n \quad (2)$$

확률분포.

C_1일 확률

$y_n = \sigma(\theta^T x_n) = P(C_1 \mid x_n) = 0$

[0,1] 실수

$t_n = \begin{cases} +1 & C_1 \\ 0 & C_2 \end{cases}$

$1 - y_n = P(C_2 \mid x_n)$

실수 정수

질문: 이거랑 뻰하대랑 관계 없지 않나?

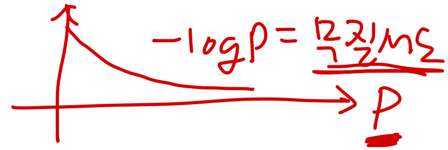
Discrete random variable: X

Probability mass function: $p(x)$

Entropy = 평균 정보량 = 평균 무질서도

$$H(p) = \mathrm{E}[-\log p] = -\sum p(x) \log p(x) \quad (3)$$

p↑ 무질서도↓, p↓ 무질서도↑

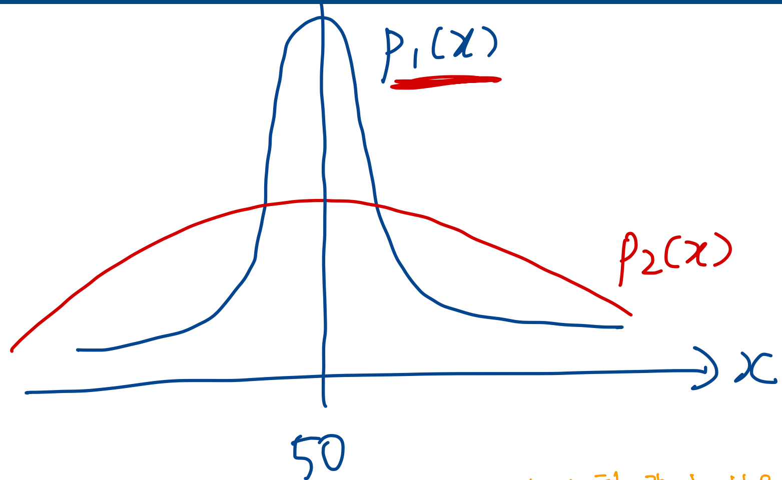


Entropy H(p) p: 확률분포.

평균

$$= E[-\log p]$$

$$= \sum_x -\log p(x) \cdot p(x)$$



$p_1(x) \rightsquigarrow$ 49, 47, 51, 53, 48, 52, ...

$p_2(x) \rightsquigarrow$ 31, 62, 53, 34, 47, ~

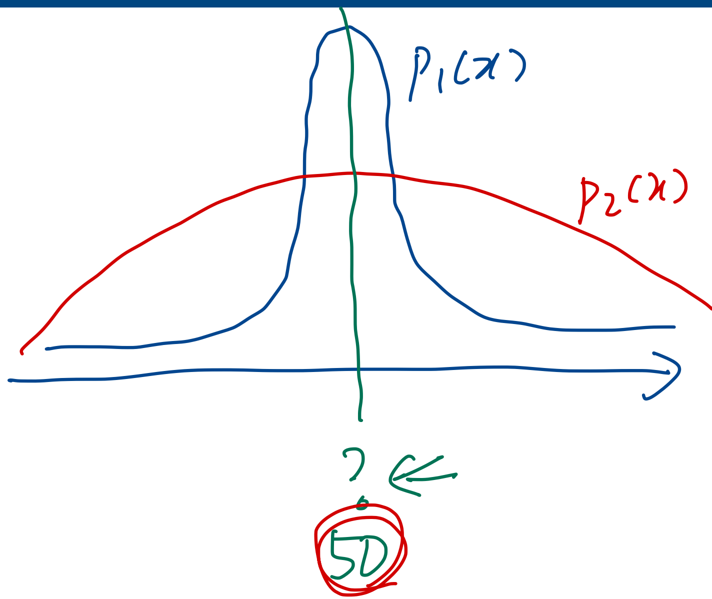
예측가능한 값이 나온다.

'하나의 샘플이 가지고 있는 정보가 작다.' ← 뻔하다 = 무작위성 ↓ = 엔트로피 ↓

평균무질서도

하나하나의 샘플이 가진 정보가 많다. ← 덜 뻔하다 = 무작위성 ↑ = 엔트로피 ↑

평균무질서도

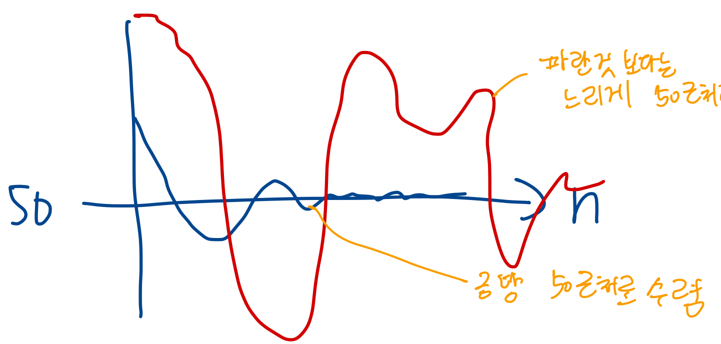


$p_1(x) \sim 51, 49, 48, 52, 53, \ldots$

$\bar{x} = \frac{1}{N}\sum x_n$

$p_2(x) \sim 31, 47, 59, 62, 38, \ldots$

$\bar{x} = \frac{1}{N}\sum x_n$



하나의 샘플이 추가될때마다 평균을 측정하는데 기여하는 정도 = 정보량.

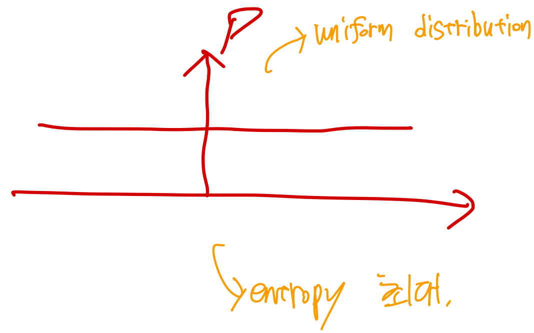
Entropy는 확률분포 한개에 대해서 말하는것.

Entropy와 불확실성, 그리고 정보량

Entropy가 최대인 확률 분포

- Discrete: uniform distribution
- Continuous: Gaussian distribution

질문: 최소 아닌가?



Cross-entropy and relative entropy

두개의 확률분포

Cross-entropy

p, q

$$\text{Entropy} = E_p[-\log p] = \sum_x -p(x)\log p(x)$$

$$H(p, q) = -E_p[\log q] = -\sum p(x)\log q(x) \quad (4)$$

$$\neq H(q, p)$$

Relative entropy

KL divergence

p, q

$$\mathcal{D}_{\mathrm{KL}}(p \,||\, q) = E_p\left[\log \boxed{\frac{p}{q}}\right] = -\sum p(x) \log \frac{p(x)}{q(x)} \quad (5)$$

$$= -\sum p(x)\{\log p(x) - \log q(x)\}$$

두 분포가 같을 때

$$\mathcal{D}_{\mathrm{KL}}(p \,||\, q) = 0 \iff p(x) = q(x) \quad \forall x \quad (6)$$

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

0보다 작을수도 있지 않나?

∴ 차이의 정도 p, q

Cross-entropy and relative entropy

고정

Cross-entropy = entropy + relative entropy

측정.

이거는 고정

$$\rightarrow H(p, q) = H(p) + \mathcal{D}_{\mathrm{KL}}(p \,||\, q) \tag{7}$$

$$-\sum p \log q = -\sum p \log p + \sum p \log p - p \log q$$

min cross-entropy $\Longleftrightarrow$ min relative entropy

$$D_{KL}(p||q) = E_p\left[\log p/q\right] = \sum p(x) \log \frac{p(x)}{q(x)}$$

$$H(p,q) = H(p) + D_{KL}(p||q)$$

$$-\sum p(x) \log q(x) = -\sum p(x) \log p(x) + \sum p(x) \log \frac{p(x)}{q(x)}$$

$$= \cancel{-\sum p(x) \log p(x)} + \cancel{\sum p(x) \log p(x)} - \sum p(x) \log q(x)$$

NN → softmax layer → 예측 확률 : 각 class에 속할 확률.

$\Updownarrow$ $\boxed{\text{cross-entropy}}$

참값

타겟 1, 0, 0, 0

유사하게.