

Preclass 06: Information Theory

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

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

Logistic regression

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

두개의 확률 분포: 2노드

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

Taking the gradient of the error function, we obtain

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

$y_n = \sigma(\mathbf{\Theta}^T \mathbf{x}_n) = P(C_1 | \mathbf{x}_n)$ [0,1] 실수

$1 - y_n = P(C_2 | \mathbf{x}_n)$

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

실수, 정수

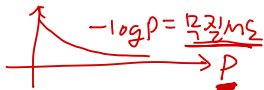
Entropy

Discrete random variable: X

Probability mass function: $p(x)$

Entropy = 평균 정보량 = 평균 무질서도

$p \uparrow$ 무질서도 $\downarrow$, $p \downarrow$ 무질서도 $\uparrow$

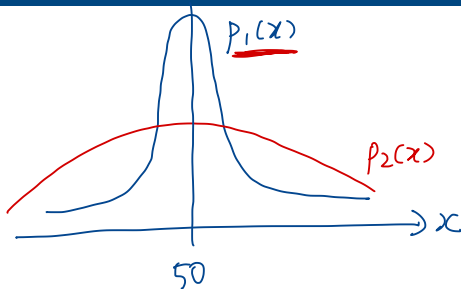


$$H(p) = \mathbb{E}[-\log p] = -\sum p(x) \log p(x) \quad (3)$$

Entropy $H(p)$ p : 확률 분포.

$$= \mathbb{E}[-\log p]$$

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

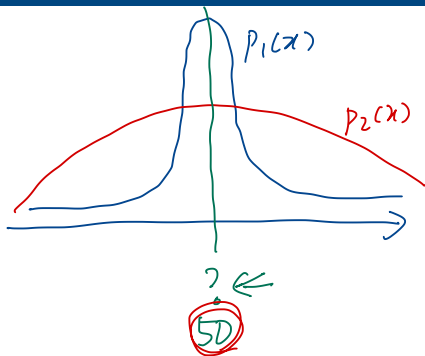


$$p_1(x) \sim \underbrace{49}, \underbrace{47}, \underbrace{51}, \underbrace{53}, \underbrace{48}, \underbrace{52}, \dots$$

$$p_2(x) \sim 31, 62, 53, 34, 47, \dots$$

← 하나의 샘플이 가지고 있는 정보量少다. 한번 → 무작위성 ↓ = 인트루피 ↓
평균무질서도

← 하나의 샘플이 가진 정보多다. 몇번 → 무작위성 ↑ = 인트루피 ↑
평균무질서도



$$p_1(x) \sim \underline{51}, \underline{49}, \underline{48}, \underline{52}, \underline{53}, \dots$$

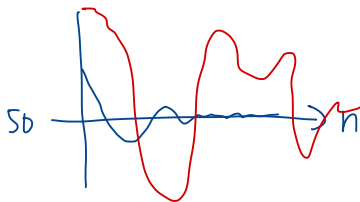
$$\bar{x} = \frac{1}{N} \sum x_n \quad \uparrow$$

$$p_2(x) \sim \underline{31}, \underline{47}, \underline{59}, \underline{62}, \underline{38}, \dots$$

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

해설의 샘플이 작아질수록
평균을 측정하는데

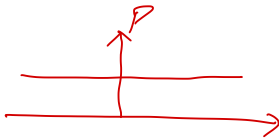
기여하는 차분 = 샘플량.



Entropy와 불확실성, 그리고 정보량

Entropy가 최대인 확률 분포

- Discrete: uniform distribution
- Continuous: Gaussian distribution



Cross-entropy and relative entropy

Cross-entropy

p, q

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

$$\begin{aligned} H(p, q) &= -E_p[\log q] = -\sum p(x) \log q(x) \quad (4) \\ &\neq H(q, p) \end{aligned}$$

0 Relative entropy

KL divergence

p, q

$$\begin{aligned} \mathcal{D}_{\text{KL}}(\underline{p} \parallel q) &= E_p \left[\log \frac{p}{q} \right] = -\sum p(x) \log \frac{p(x)}{q(x)} \quad (5) \\ &= -\sum p(x) \{ \log p(x) - \log q(x) \} \end{aligned}$$

두 분포가 같을 때

$$\underline{\underline{\mathcal{D}_{\text{KL}}(p \parallel q) = 0}} \iff \underline{\underline{p(x) = q(x) \quad \forall x}} \quad (6)$$

$$\mathcal{D}_{\text{KL}}(p \parallel q) \geq 0$$

$\simeq$ 제너리의 정제 p, q

Cross-entropy and relative entropy

Cross-entropy = entropy + relative entropy

$\approx \mathcal{H}_0$.

$$\rightarrow H(p, q) = H(p) + \mathcal{D}_{\text{KL}}(p \parallel q) \quad (7)$$

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

min cross-entropy $\Leftrightarrow$ min relative entropy

NN

→ Softmax
layer

예측
확률

: 각 class에 속할 확률.



Cross-entropy

참가
값

태그

1, 0, 0, 0

→
구사함.