

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

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

Taking the gradient of the error function, we obtain

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

Entropy

Discrete random variable: X

Probability mass function: $p(x)$

Entropy = 평균 정보량 = 평균 무질서도

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

Entropy가 최대인 확률 분포

- Discrete: uniform distribution
- Continuous: Gaussian distribution

Cross-entropy and relative entropy

Cross-entropy

$$H(p, q) = -\mathbb{E}_p[\log q] \quad (4)$$

Relative entropy

$$\mathcal{D}_{\text{KL}}(p \parallel q) = \mathbb{E}_p \left[\log \frac{p}{q} \right] \quad (5)$$

두 분포가 같을 때

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

Cross-entropy and relative entropy

Cross-entropy = entropy + relative entropy

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