N / A / N / O / D / E / G / R / E / E

지도학습 알고리즘

07. Information Theory

Logistic regression

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

P.q.

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

min.

예측. $y_n = P(C_1 | x_n)$

$1 - y_n = P(C_2 | x_n)$.

t_n = C_1이면 +1, C_2이면 0.

$1 - t_n$ = C_1이면 0, C_2면 +1.

- Taking the gradient of the error function, we obtain

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

참 확률분포.

예측 확률분포.

$t_n = 1$인 경우, $P(C_1 | x_n) = 1$

$t_n = 0$, $P(C_1 | x_n) = 0$

Entropy

- Discrete random variable

$$X$$

- Probability mass function

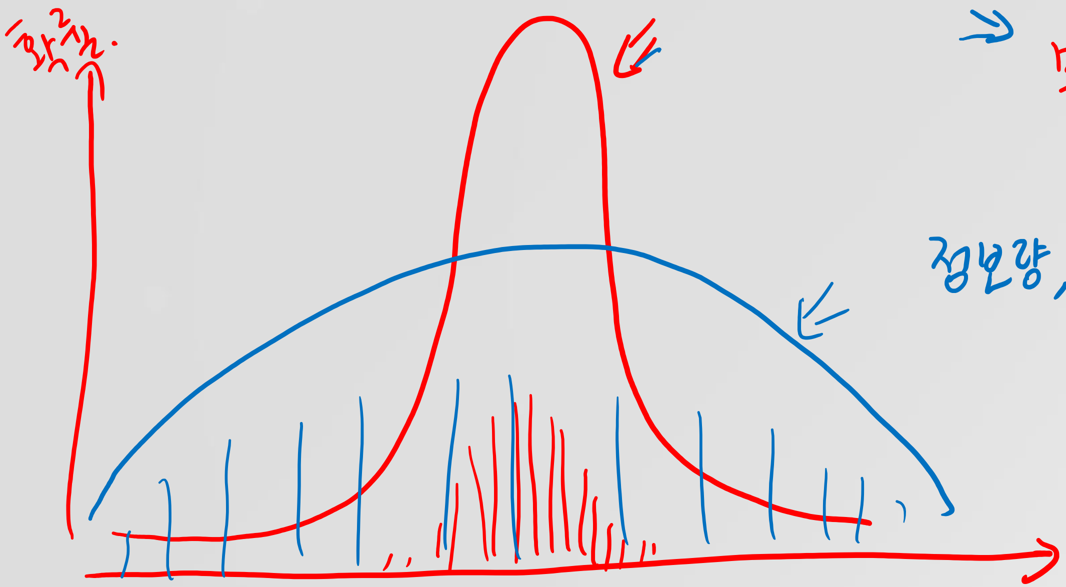
$$p(x)$$

- Entropy = 평균 정보량 평균 무질서도.

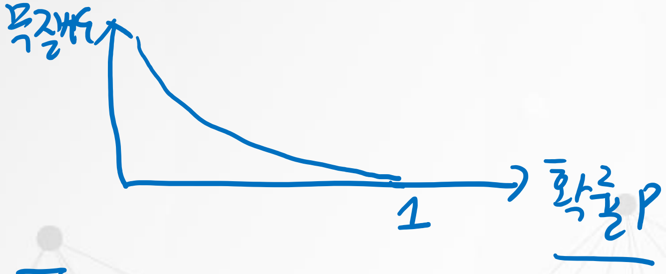
$$H(p) = \mathrm{E}[-\log p] = \boxed{-\sum p(x)\log p(x)}$$

Entropy와 불확실성, 그리고 정보량

entropy = 무질서도, 얼마나 무작위하냐.



→ Entropy 무질서함 : 빨간색 분포 < 파란색 분포.



정보량, 무질서도. $-\log p$

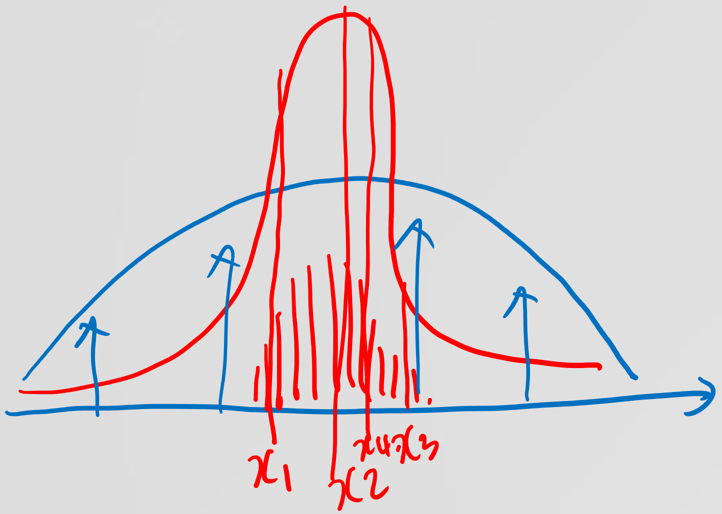
Entropy 평균무질서도. $E[-\log p] = \sum_x -p(x)\log p(x)$

$X \sim p(x)$

기대값. $E[x] = \sum x p(x)$

Entropy와 불확실성, 그리고 정보량

무질서도 ∝ 정보량



각각의 분포에서 샘플을 추출 → 각분포의 평균을 측정.

빨간색 분포에서 추출된 sample → 1개, 2, 3, 4, ...
↑ ↑ ↑ 기여하는 바.

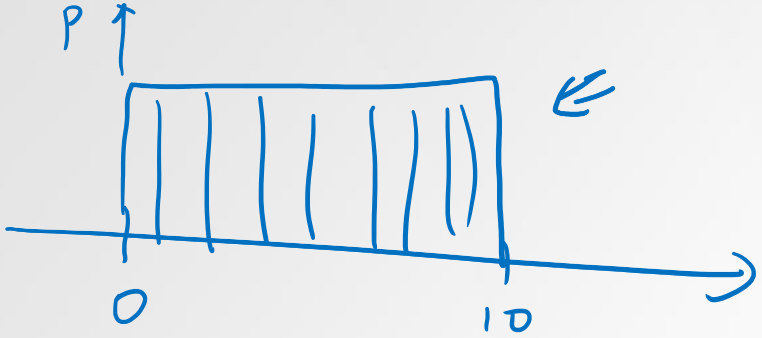
∧

파란색 분포 " → 1개, 2, 3, 4, ...
↑ ↑ ↑ 기여하는 바. ⇐
정보량.

Entropy와 불확실성, 그리고 정보량

Entropy가 최대인 확률 분포

- Discrete: uniform distribution
- Continuous: Gaussian distribution



신경망 → 'cross-entropy'

softmax. → 확률값.

expectation-maximization

Variational approximation ← relative entropy

통계적 추론 ← efficient estimator

Cramer-Rao lower bound

Cross-entropy and relative entropy

- Cross-entropy

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

Entropy $E[-\log p(x)] = -\sum p(x) \log p(x)$

cross-entropy. $H(p,q) = -E_p[\log q]$

$= -\sum p(x) \log q(x)$.

- Relative entropy , KL divergence.

$$\mathcal{D}_{\mathrm{KL}}(p||q) = \mathrm{E}_p\left[\log \frac{p}{q}\right] = \sum p(x) \log \frac{p(x)}{q(x)} \boxed{\geq 0}$$

1

- 두 분포가 같을 때, relative entropy = 0

$$\mathcal{D}_{\mathrm{KL}}(p||q) = 0 \Longleftrightarrow p(x) = q(x)$$

작을수록 비슷한값.

두 확률 분포의 similarity

p, q_1, q_2 $D_{KL}(q_1||p)$

$D_{KL}(q_2||p)$

Cross-entropy and relative entropy

- Entropy, cross-entropy and relative entropy

$$\underbrace{H(p,q)}_{\text{cross entropy } p,q} = \underbrace{H(p)}_{p\text{의 entropy}} + \underbrace{D_{\mathrm{KL}}(p||q)}_{q\text{에대한 } p\text{의 relative-entropy}} \geq 0.$$

$= 0$ p=q 동일할때.

$\min H(p,q) \Rightarrow \min D_{KL}(p||q)$

$\Rightarrow$ 우리가원하는건 $p=q$ 같게. 최대한 유사하게.