

기계학습 6주차 과제

12191656 이채연

HomeWorks #1

HomeWorks #1

- When $X \in \{0,1\}$, $p(X = 1) = \theta$ and $p(X = 0) = 1 - \theta$. Find $\mathbb{H}(X)$.

$$\begin{aligned} H(X) &= - \sum_{k=1}^K P(X=k) \log_2 P(X=k) \\ &= - \left[\frac{P(X=1)}{\theta} \log_2 P(X=1) + \frac{P(X=0)}{1-\theta} \log_2 P(X=0) \right] \quad (P(X=1)=\theta, P(X=0)=1-\theta \text{ 이므로}) \\ &= - \left[\theta \log_2 \theta + (1-\theta) \log_2 (1-\theta) \right] \end{aligned}$$

결과적으로 이러한 binary random variable에 대한 entropy는 cross entropy 형태로 표현이 된다.

HomeWorks #2

Explain the difference between Entropy and KL divergence. Provide the probability definitions for both.

- Difference between Entropy and KL divergence

Entropy 같은 경우에는 하나의 random variable에 대한 uncertainty(불확실성)만 내조하였다면, KL divergence라는 것은 하나가 아니라 두 개의 probability distribution의 dissimilarity(비유사성), 즉 similarity가 얼마나 없는지를 내조하는 지표이다.

- Probability definition of Entropy

Random variable X 에 대한 Entropy 정의는 다음과 같다. 여기서 random variable X 에 대한 distribution을 p 라고 표시하고, X 의 state가 K 개의 state가 있다고 가정한다.

$$H(X) \triangleq - \sum_{k=1}^K p(X=k) \log_2 p(X=k)$$

- Probability definition of KL divergence

두 probability distribution p 와 q 의 KL divergence 정의는 다음과 같다.

$$KL(p \parallel q) \triangleq \sum_{k=1}^K p_k \log \frac{p_k}{q_k} = \sum_{k=1}^K p_k \log p_k - \sum_{k=1}^K p_k \log q_k = -H(p) + H(p, q)$$

여기서 $H(p)$ 는 "regular"(일반적인) entropy이고, $H(p, q)$ 는 cross entropy이다. p 가 알고 싶어하는 distribution이고, q 가 알고있는 distribution이라고 할 때, 알고있는 distribution q 를 가지고 p 를 modeling할 때 얼마나 많은 데이터(평균 bit 수)가 필요한지를 나타내는 것이 cross entropy이다. 참고로 두 probability distribution p 와 q 의 cross entropy의 정의는 다음과 같다.

$$H(p, q) = \sum_{k=1}^K p_k \log q_k$$