

KL-Divergence에서 Maximum Log Likelihood까지...

두 distribution간의 거리

$$D[P_{data} || P_{\theta}] \triangleq E_{x \sim P_{data}} \left[\log \frac{p_{data}(x)}{p_{\theta}(x)} \right] \quad \text{x가 P_data 분포를 따를 때, 괄호 안의 식의 기댓값}$$

가능한 모든 x에 대해 더한다
x가 continuous domain에 있는 경우,
summation 대신 integral

$$\triangleq \sum_x p_{data}(x) \left[\log \frac{p_{data}(x)}{p_{\theta}(x)} \right] \quad \text{P_data에서의 likelihood에 해당하는 가중치를 적용해서}$$

$$= \underbrace{E_{x \sim P_{data}} [\log p_{data}(x)]}_{\text{P_data의 negative entropy}} - \underbrace{E_{x \sim P_{data}} [\log p_{\theta}(x)]}_{\text{P_data와 P_theta의 cross entropy}}$$

P_theta에 대해 상수항

두 distribution간의 거리를 최소화하는 P_theta를 구하자

min문제를 max문제로 바꾸고 negative 부호 제거

$$\arg \min_{p_{\theta}} D[P_{data} || P_{\theta}] = \arg \max_{p_{\theta}} E_{x \sim P_{data}} [\log p_{\theta}(x)] \quad \text{P_data와 P_theta의 cross entropy를 최소화하는 P_theta를 구하자}$$

$$= \arg \max_{p_{\theta}} \frac{1}{|D|} \sum_{x \in D} \log p(x)$$

전체 dataset 크기

True P_data의 분포를 알 수 없으므로 P_data에서 샘플링 됐다고 가정한 D=dataset를 이용하여 Monte Carlo Estimation을 수행