

Kullback-Leibler Divergence & Jensen-Shannon Divergence

<사용 목적> : 서로 다른 확률 분포의 차이를 측정하기 위하여

1. 특정 dataset의 분포를 추정하기 위해 그 집단을 대표하는 sample을 뽑아 관찰

-> sample을 통해 데이터의 분포를 추정했을 때 추정한 분포가 얼마나 잘 추정한 것인지 측정할 때 사용

2. 복잡한 확률 분포를 비교적 적은 파라미터를 갖는 Gaussian Distribution 등으로 근사하여 정보를 저장하는데 드는 비용을 크게 절감(약간의 오차는 있겠지만)하려할 때, 가장 오차가 적은 model로 근사하기 위해 사용

<상황 가정>

- 어떠한 알파벳이 뽑힐 확률이 각각 있음(1/9, 2/9, 1/3, 2/9, 1/9)

-> 표로 저장하고 나타내는 것은 정보를 저장하는데 많은 용량이 필요, 수학적으로 다루기도 까다로움

▶ 확률분포근사

- uniform Distribution, Binomial Distribution 등으로 근사시킬 수 있음.

-> 각 확률분포는 실제 상황에서 유도된 것이지만, 해당 확률들과 비슷하게 맞추어주기 위해 임의로 가정해서 사용할 수 있음

<Entropy> : 평균 정보량

- 어떠한 사건 A가 있고 각 item의 확률이 있다. 임의로 어떠한 item을 뽑았을 때, 해당 item인지를 알기 위해 물어보아야 하는 질문의 평균(=답아야 하는 정보의 양의 평균)(실제 각 item 모든 경우를 생각)

- 질문의 평균을 낮추려면 확률이 0.5씩 갈라지도록 질문하는 것이 좋음.

- 그렇게 했을 때, 각 item의 질문 개수 : $\log p^{-1}$

=> Entropy : $\sum_i p_i \log p^{-1}$

- 확률분포가 무작위하지 않을수록 entropy 값이 작아지므로 dataset의 무작위성을 판단할 때에도 사용됨

- entropy가 낮다면, 어떠한 사건에서 item이 왔을 때, 평균적으로 적은 질문으로 무엇인지 알아냄. 평균적으로 적은 정보의 양을 담고있음.(질문의 양이 많다면 담고 있는 정보의 양도 많음)

<Cross Entropy>

- 어떠한 사건 A가 있고 해당 확률 분포가 $p(x)$ 일 때, 이를 $q(x)$ 의 확률분포라고 생각했을 때의 entropy. 원래 확률 분포와 비슷하게 생각했을 때 해당 entropy가 낮아지기 때문에, $p(x)$ 를 잘 예측했는지에 대한 평가 지표로 사용됨

=> Cross Entropy : $H(p, q) = \sum_i p_i \log q^{-1}$

<Kullback-Leibler Divergence>

- q라는 분포를 사용했을 때의 p의 entropy와 실제 p의 entropy의 차이를 사용하여 두 분포간의 차이를 측정(p와 q의 분포가 비슷할수록 cross-entropy가 작아지기 때문에 사용 가능)

$$KL(p||q) = H(p,q) - H(p)$$

▶ cross entropy로부터 KL divergence 유도

$$H(p,q) = - \sum_i p_i \log q_i$$

$$= - \left(\sum_i p_i \log p_i + \sum_i q_i \log \frac{q_i}{p_i} \right)$$

$$= H(p) + \sum_i p_i \log \frac{p_i}{q_i}$$

$$\Rightarrow \sum_i p_i \log \frac{p_i}{q_i} : \text{각 item의 정보량의 차이의 평균, } KL(p,q)$$

- cross entropy를 최소화하는 것은 KL-divergence를 최소화 하는 것과 같음

▶ 특성

- $KL(p||q) \geq 0$: cross entropy의 최솟값은 entropy이기 때문
- $KL(p||q) \neq KL(q||p)$: 거리 개념이 아님.(거리는 두 항을 바꾸어도 같아야 하기 때문)

<Jensen-Shannon divergence>

- KL-divergence를 거리개념으로 사용하기 위해 도입된 개념
- $JSD(p||q) = 1/2 KL(p||M) + 1/2 KL(q||M)$
where, $M = 1/2(p+q)$

출처 : https://hyunw.kim/blog/2017/10/27/KL_divergence.html
https://hyeongminlee.github.io/post/prob002_kld_jsd/