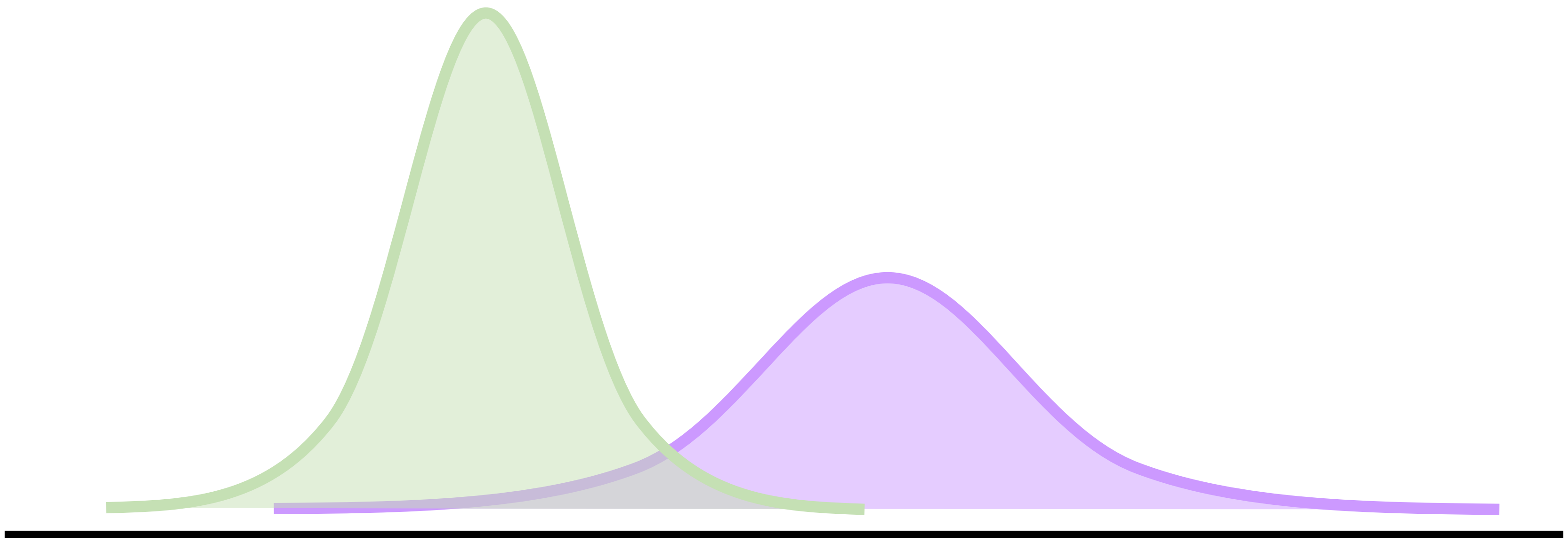
Deep Learning101

KL Divergence

확률분포 차이 구하기



신박AI

안녕하세요 여러분! 신박AI입니다

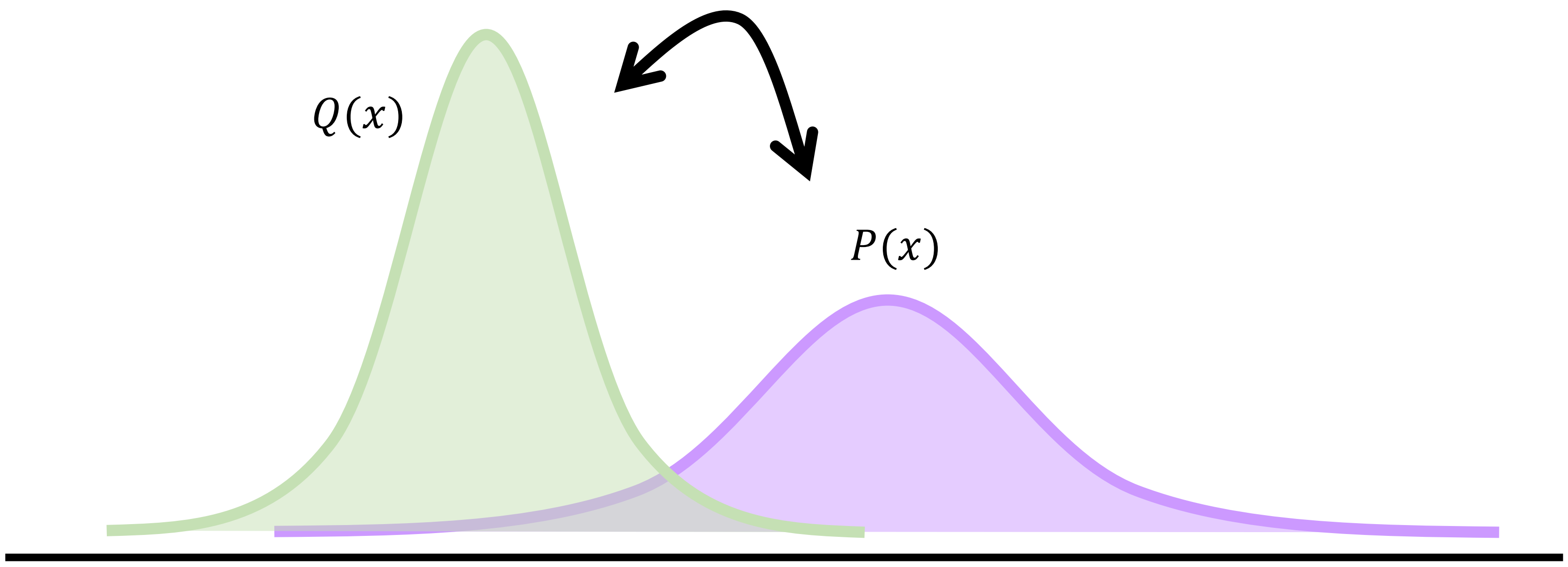
오늘의 주제는 KL Divergence입니다.

신박AI

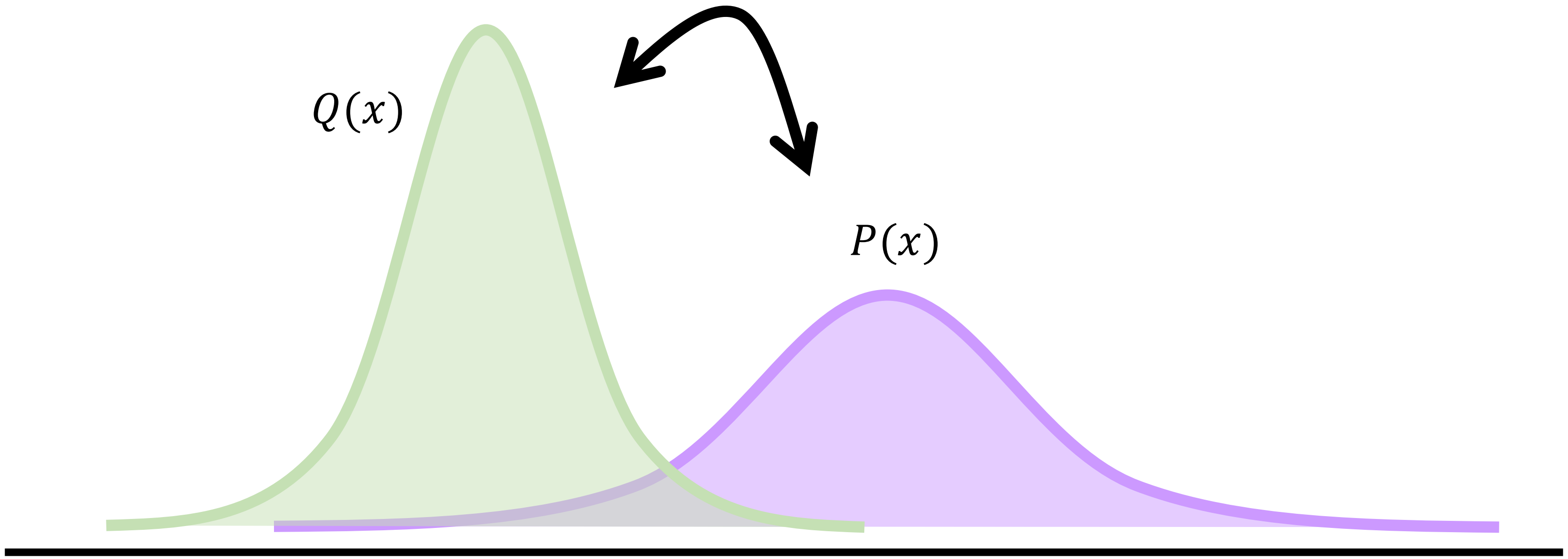
KL Divergence는 딥러닝과 머신러닝,
그리고 데이터 사이언스 분야에서는
Cross-Entropy 만큼
자주 볼 수 있는 개념입니다.

오늘 영상에서는 KL Divergence의
개념이 무엇인지, 어떻게 계산하는지,
예시와 함께 살펴보도록 하겠습니다.

KL Divergence(Kullback-Leibler Divergence)는 두 확률분포 간의 정보량의 차이를 측정하는 데 사용되는 개념입니다.



KL Divergence의 공식은 다음과 같습니다.



$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

KL Divergence의 공식은 다음과 같습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

KL Divergence의 의미를 좀 더 직관적으로 이해하기 위해 식을 다음과 같이 변형해보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \sum P(x)(\log P(x) - \log Q(x))$$

KL Divergence의 의미를 좀 더 직관적으로 이해하기 위해 식을 다음과 같이 변형해보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \sum P(x)(\log P(x) - \log Q(x))$$

$$= -\sum P(x)(\log Q(x) - \log P(x))$$

신박AI

KL Divergence의 의미를 좀 더 직관적으로 이해하기 위해 식을 다음과 같이 변형해보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \sum P(x)(\log P(x) - \log Q(x))$$

$$= -\sum P(x)(\log Q(x) - \log P(x))$$

$$= -\sum P(x)\log Q(x) - (-\sum P(x)\log P(x))$$

이렇게 보면 KL Divergence는 결국 두 분포의 Cross-Entropy에서 Entropy를 뺀 값이 됩니다.

$$
\begin{aligned}
D_{KL}(P \parallel Q) &= \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \\
&= \sum P(x)(\log P(x) - \log Q(x)) \\
&= -\sum P(x)(\log Q(x) - \log P(x)) \\
&= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}
\end{aligned}
$$

신박AI

이렇게 변형된 KL Divergence공식을 사용해 KL Divergence의 직관적 의미를 살펴보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

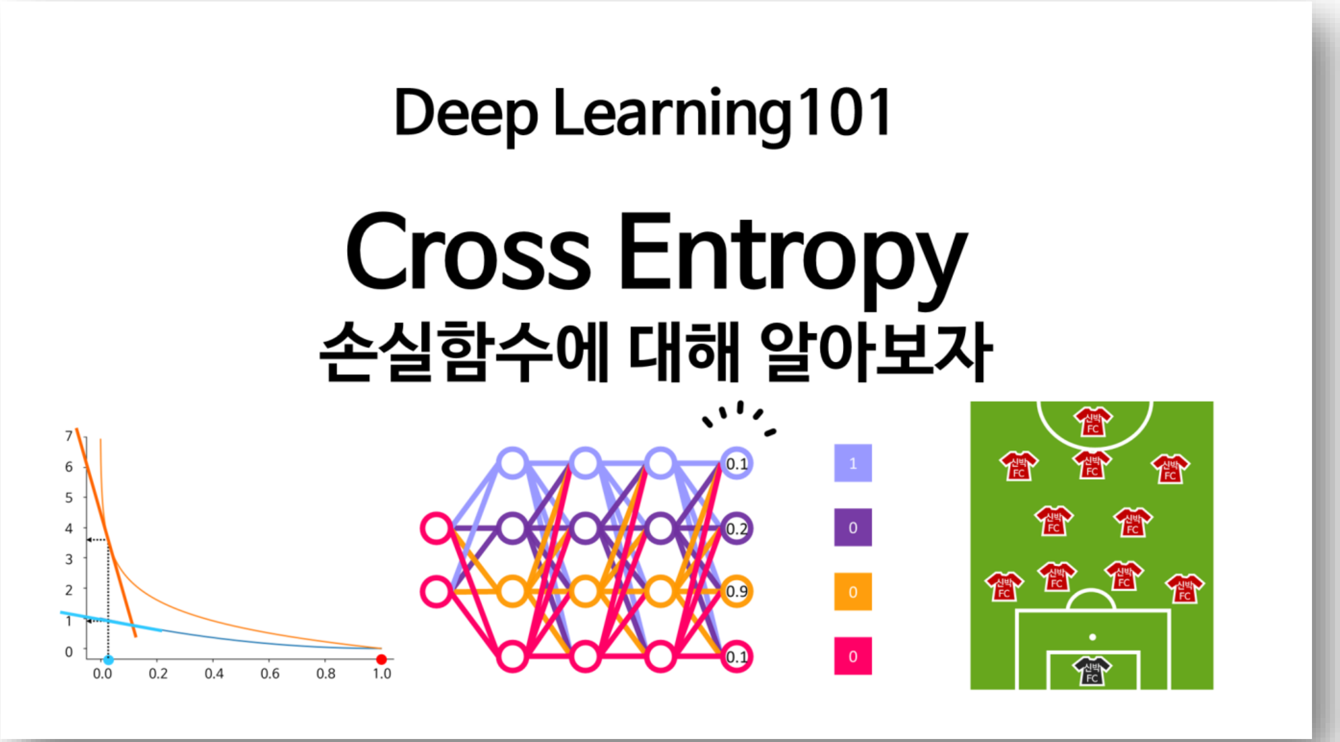
$$= \sum P(x)(\log P(x) - \log Q(x))$$

$$= -\sum P(x)(\log Q(x) - \log P(x))$$

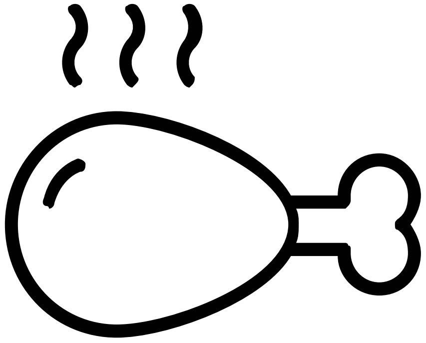
$$= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}$$

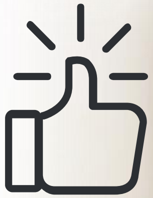
신박AI

직관적인 이해를 돕기 위해 지난 영상에서의 슬라이드를 빌려오도록 하겠습니다.

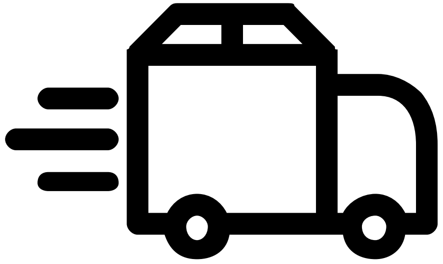


지난 시간에 Cross-Entropy의 의미를 비유적으로 맛집 리뷰와 경험치와의 차이로 말씀드린 바가 있었습니다.

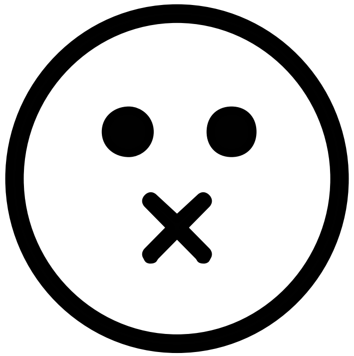




$p(x)$

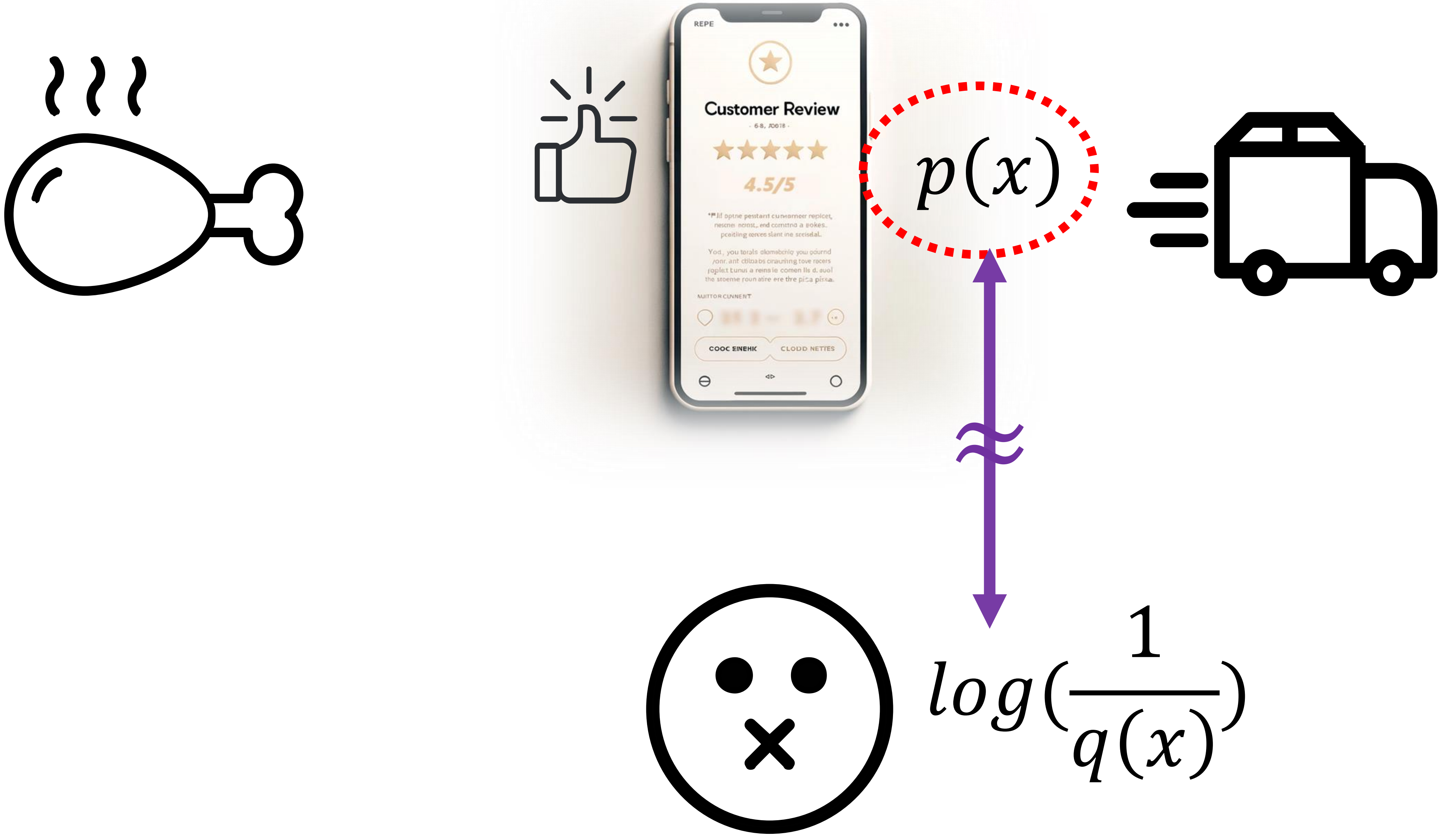


$\approx$

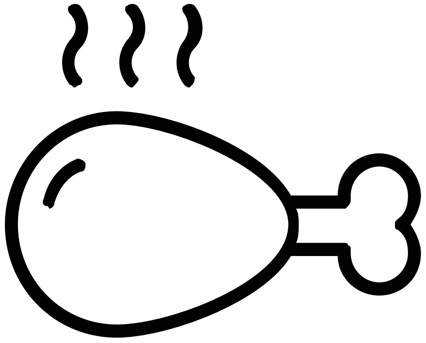


$log(\frac{1}{q(x)})$

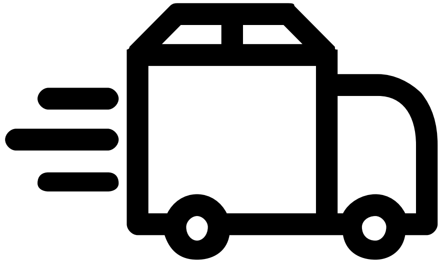
여기서 $p(x)$는 많은 사람들이 평가한 치킨집의 평점, 즉 실제 확률분포라고 할 수 있습니다.



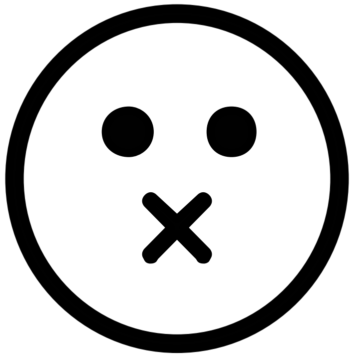
그리고 $q(x)$는 내가 평가한 점수 분포 (측정분포)라고 한다면,



$$p(x)$$

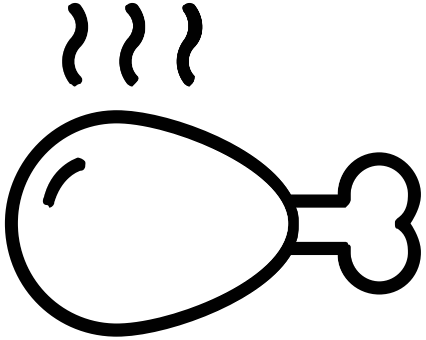


$$\not\approx$$

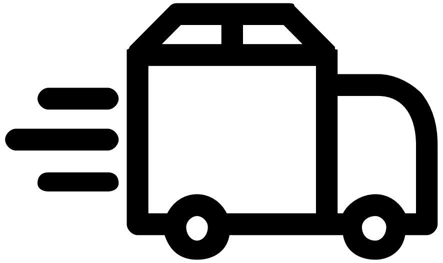


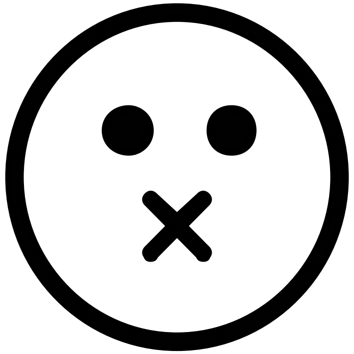
$$log(\frac{1}{q(x)})$$

측정분포 $q(x)$ 의 값이 $p(x)$에 비해 턱없이 낮을 때,



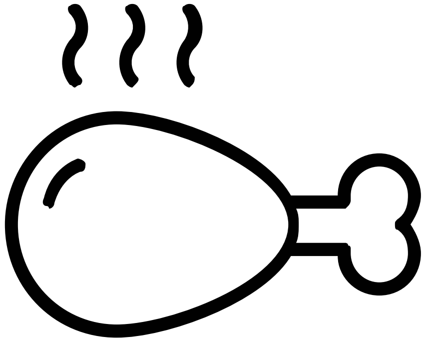
$p(x)$

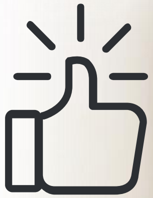




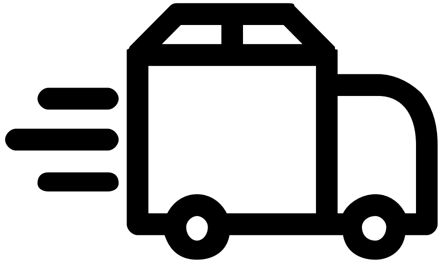
$$log\left(\frac{1}{q(x)}\right)$$

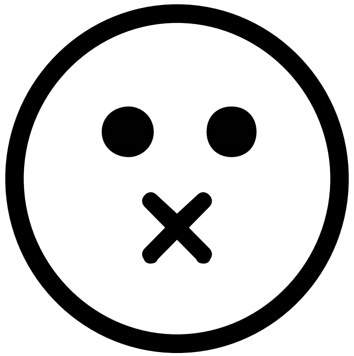
측정분포에 의한 놀람의 값이 커지게 될 것이고,





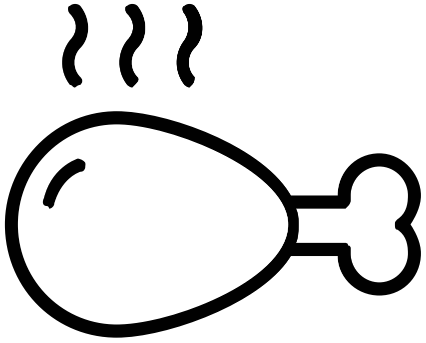
$p(x)$

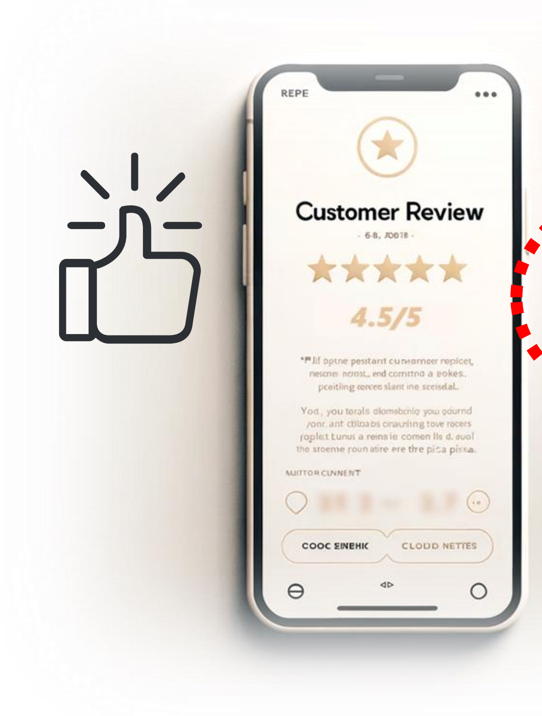




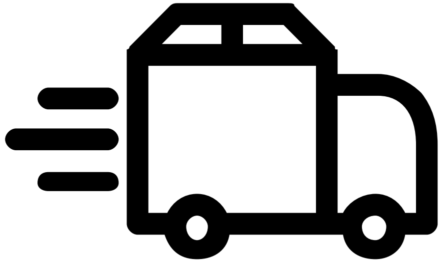
$log(\frac{1}{q(x)})$

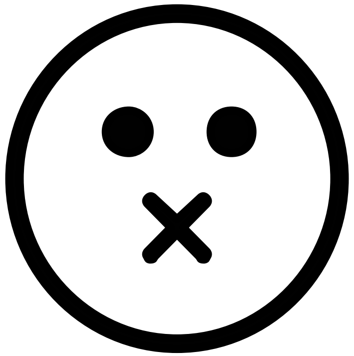
그로인해 Cross-Entropy의 값은 커지게 될 것입니다.





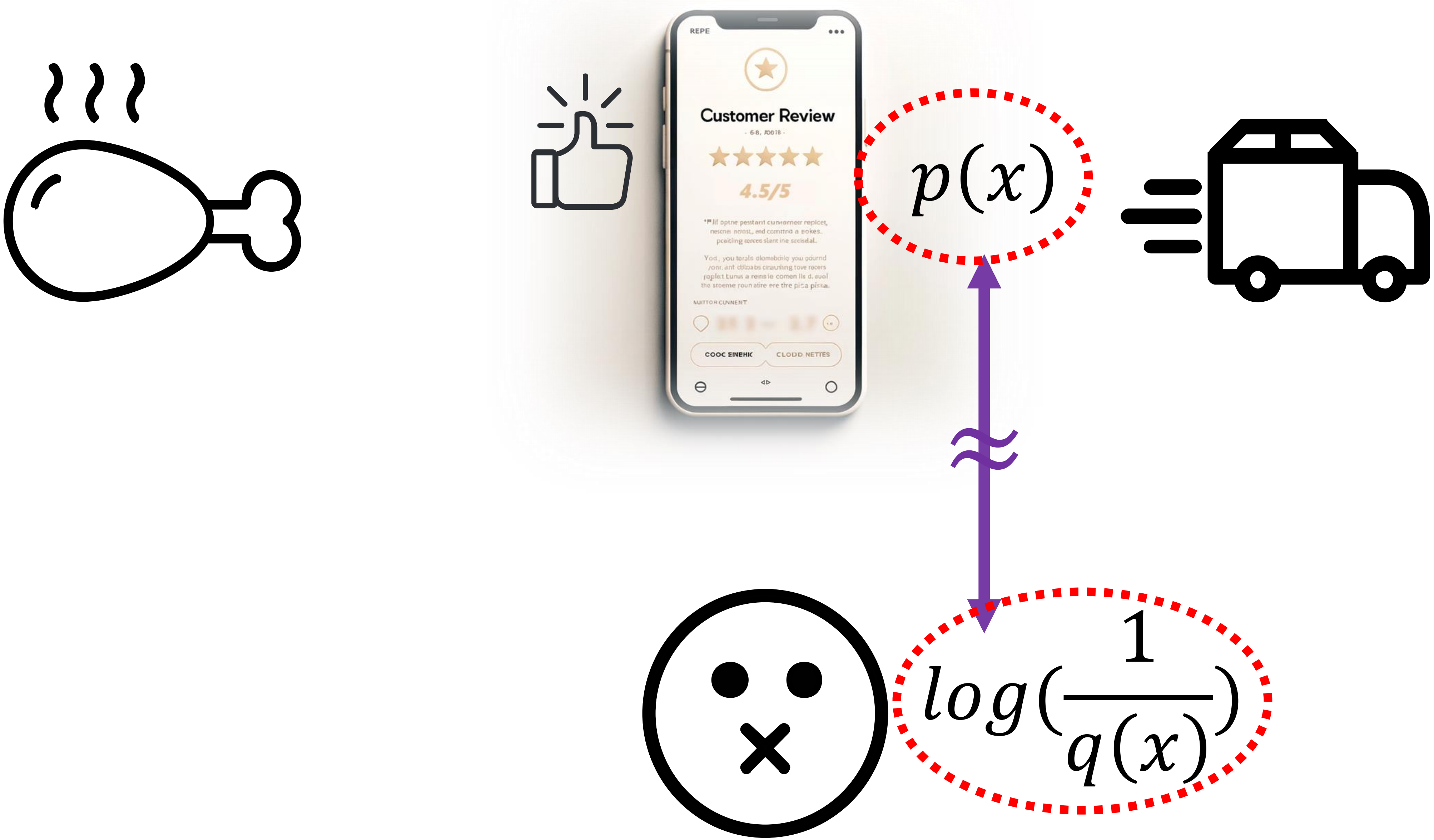
$$p(x)$$



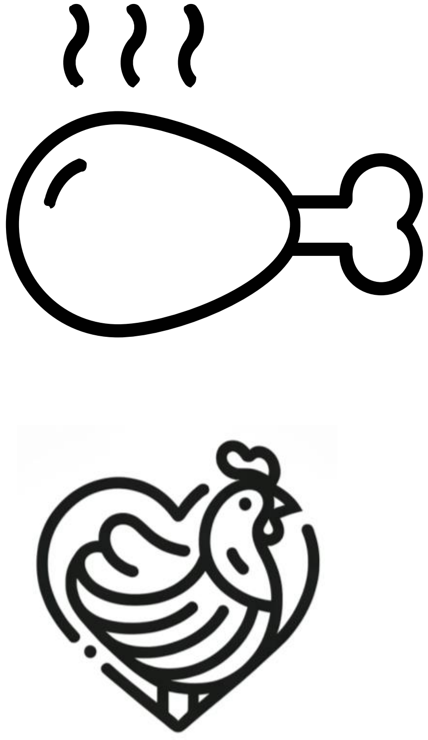


$$log(\frac{1}{q(x)})$$

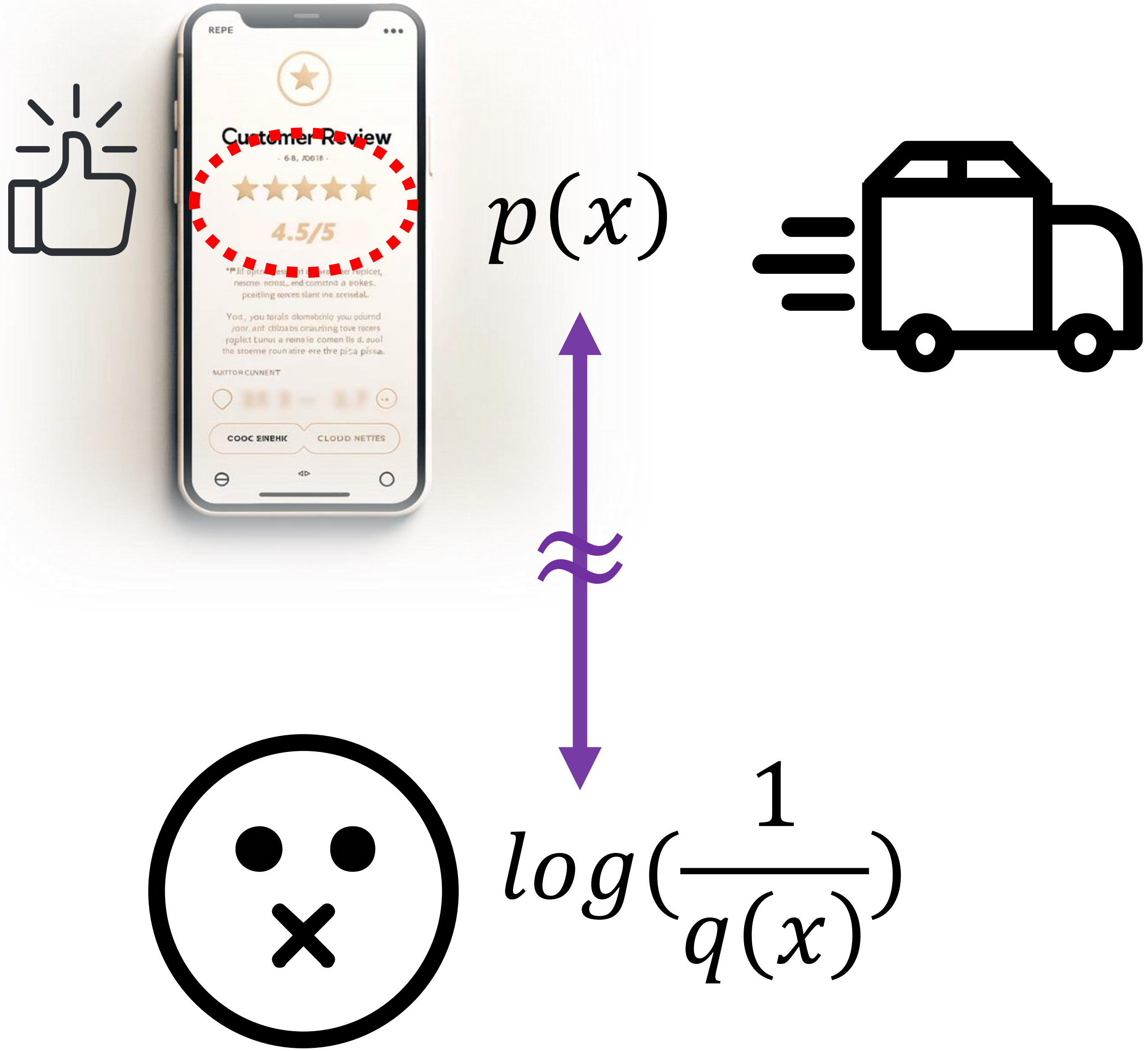
그런데 이 Cross-Entropy를 좀 더 자세히 들여다 볼 필요가 있습니다.



예를들어, 이 치킨집이 만드는 치킨이 ‘엄마 손맛 치킨’ 이라고 가정해봅시다.



엄마 손맛 치킨



‘엄마 손맛 치킨’은 그 특성상 레시피 대로 조리하여도 그 맛을 그대로 재현해내기란 힘든 법입니다.

엄마 손맛 치킨

엄마표 레시피

눈물 한 스푼
헌신 두 스푼
사랑 두 큰술
...

신박AI

말하자면, 엔트로피 (불확실성)가 높은 음식입니다.

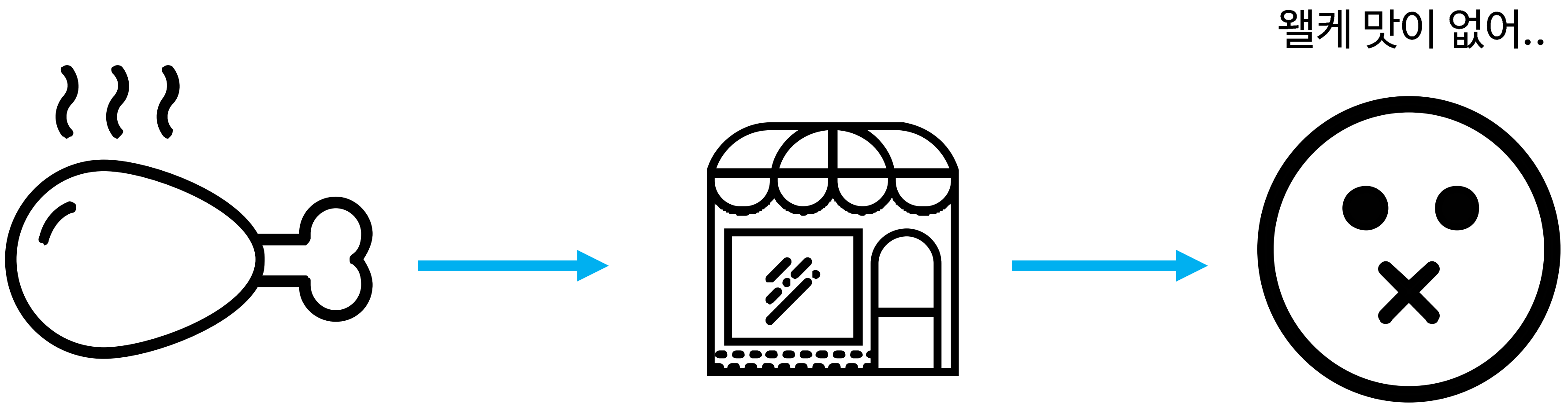
기성 제품으로는 담아내기 힘든 맛!

엄마 손맛 치킨

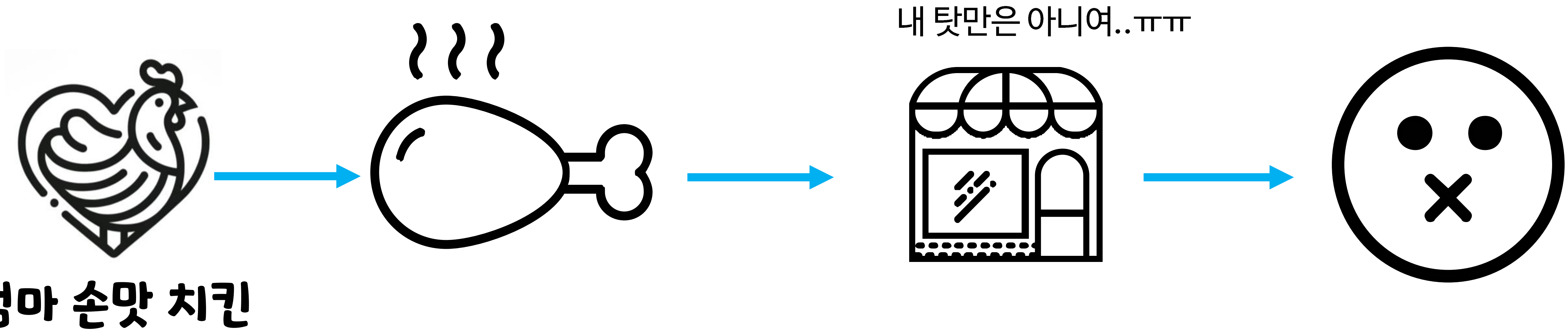
엄마표 레시피

눈물 한 스푼
헌신 두 스푼
사랑 두 큰술
...

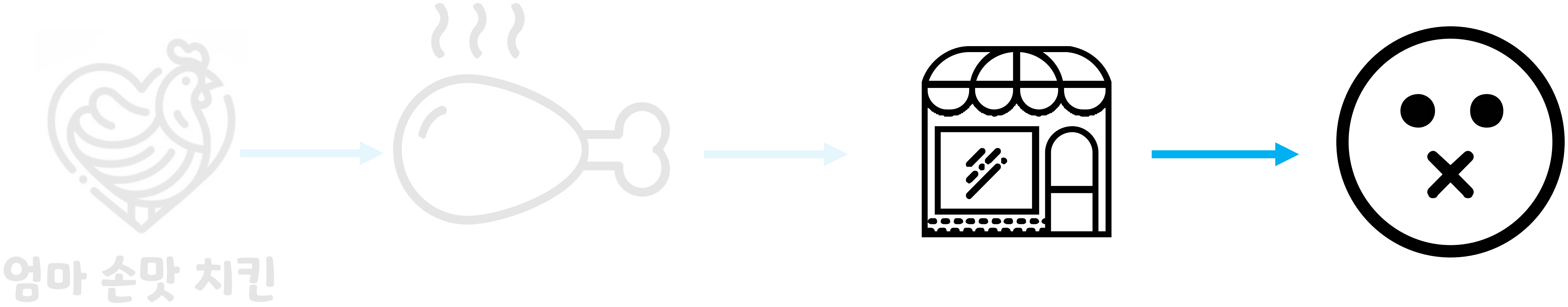
그러므로 이 치킨집의 높은 Cross-Entropy는 이 치킨집 사장님의 솜씨 때문만이라기 보다는,



엄마 손맛 치킨이라는 음식 자체의 높은 불확실성 (엔트로피) 때문이라고 볼 수도 있습니다.

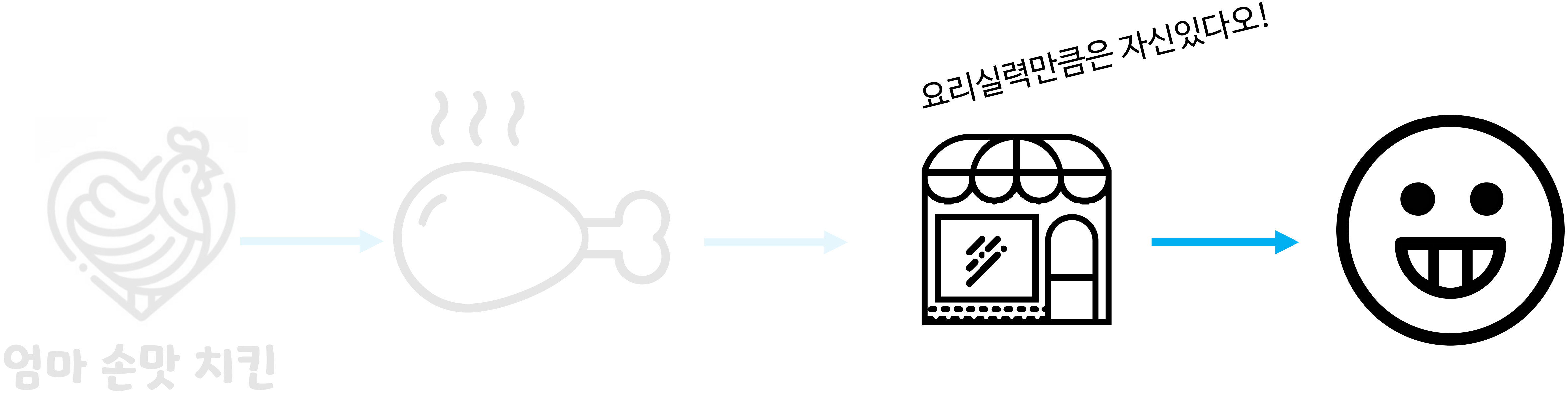


그러므로 우리는, 요리 자체의 불확실성을 제외한,



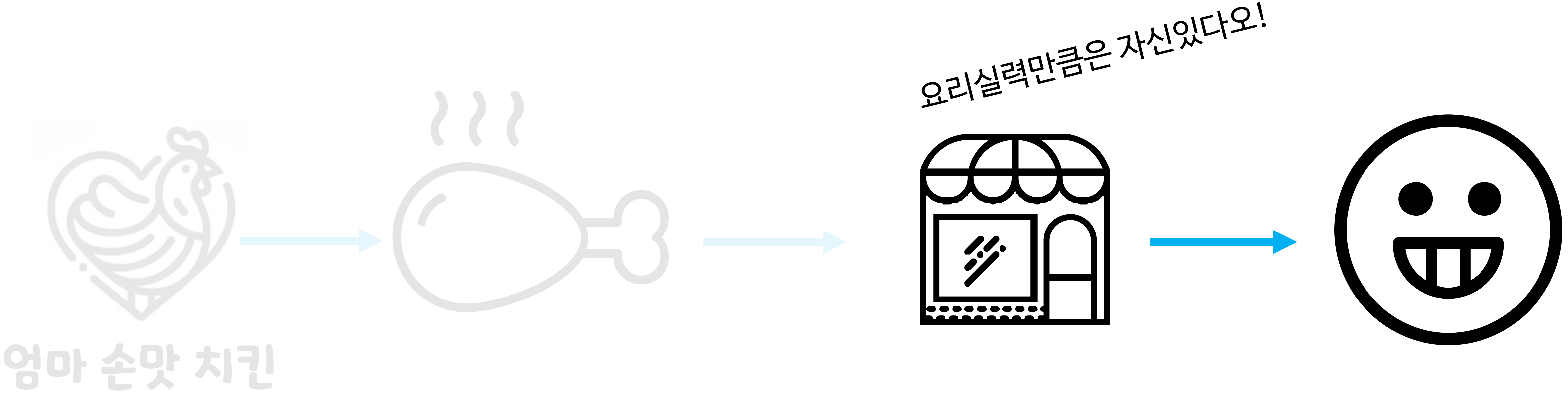
신박AI

치킨집 사장님의 순수한 요리 솜씨만을 표현하는 또 다른 척도가 있었으면 좋겠습니다.



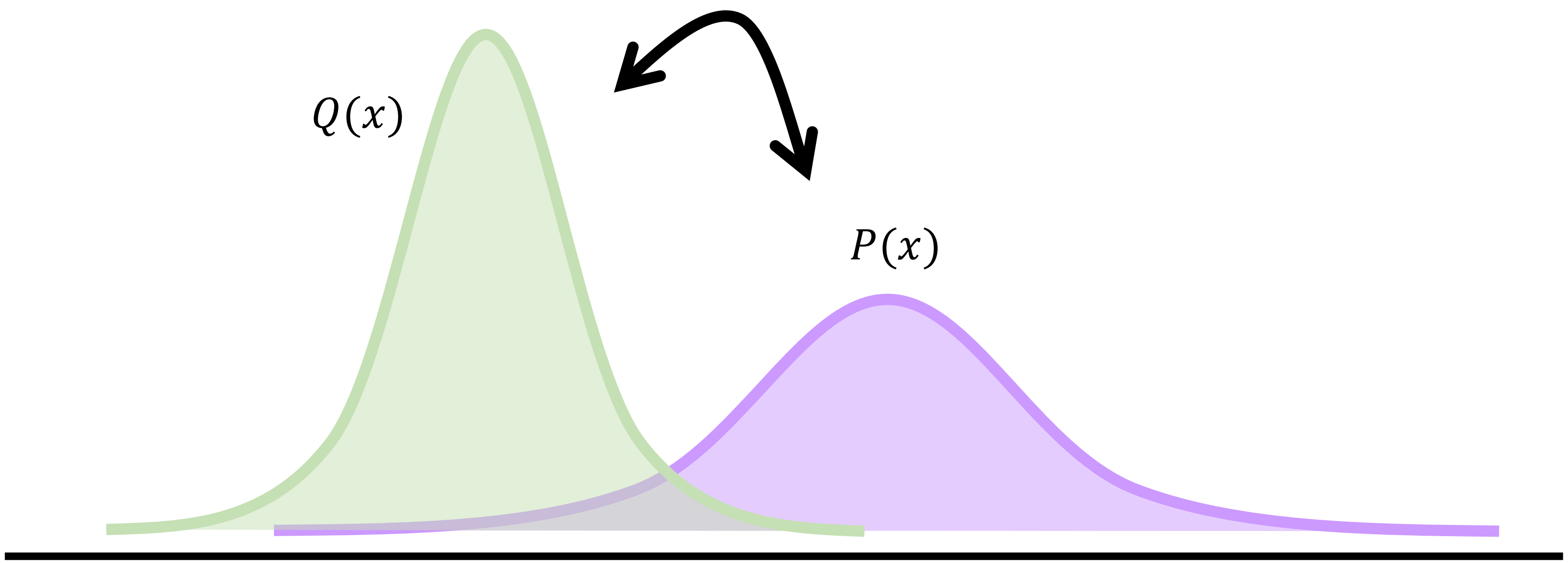
신박AI

그럴 때 쓰일 수 있는 개념이 바로 KL Divergence입니다.



신박AI

그러므로 이 KL Divergence라는 것은,

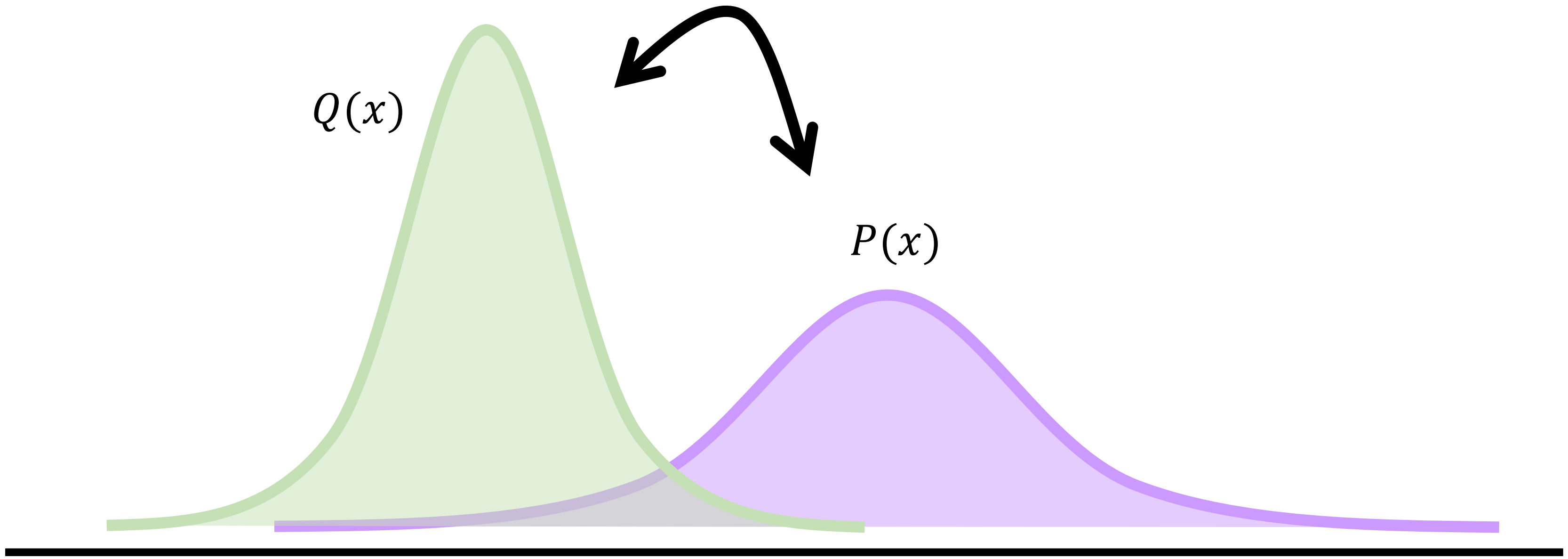


$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}$$

신박AI

Cross-Entropy 와 더불어 두 확률분포의 차이를 보여주는 척도이지만,

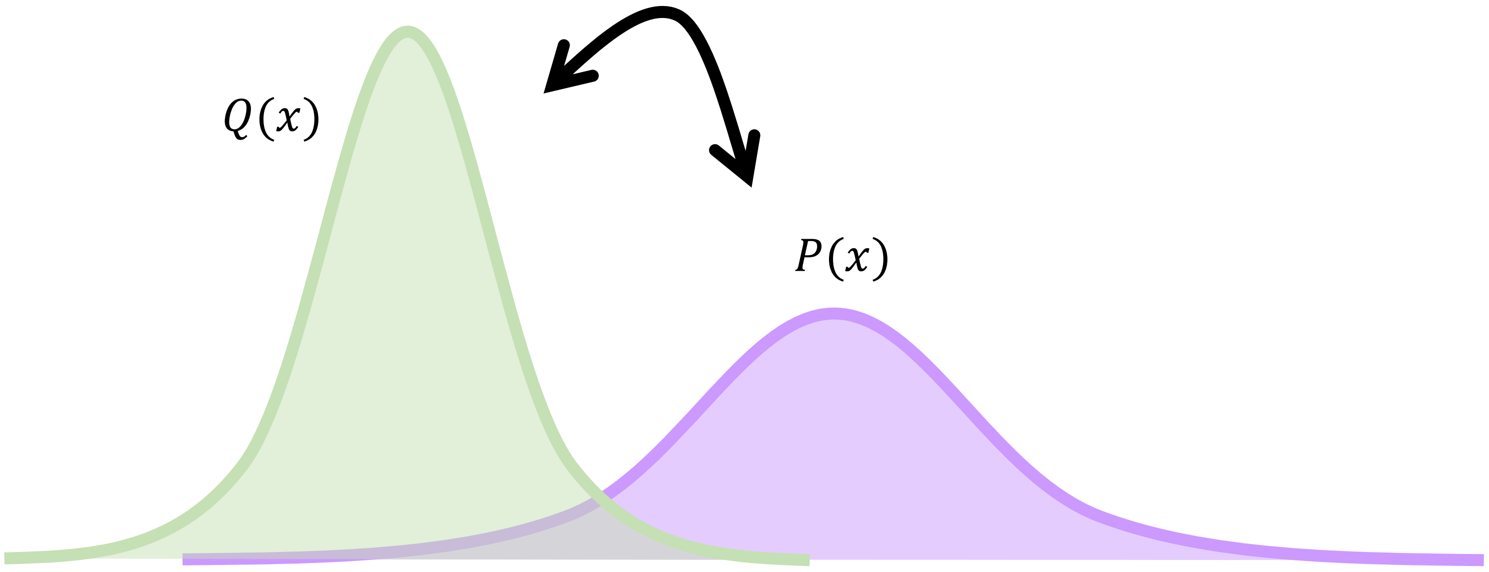


$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}$$

신박AI

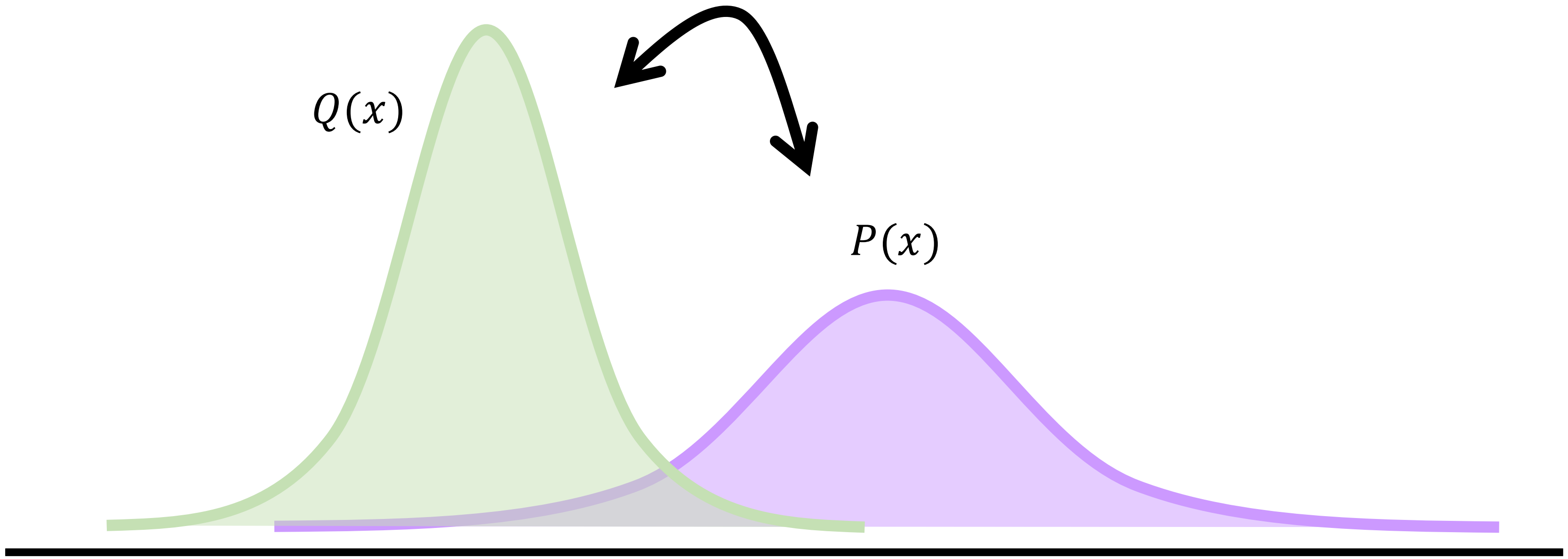
Cross-Entropy에서 기준분포 $P(x)$ 의 Entropy를 빼줌으로써,



$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}$$

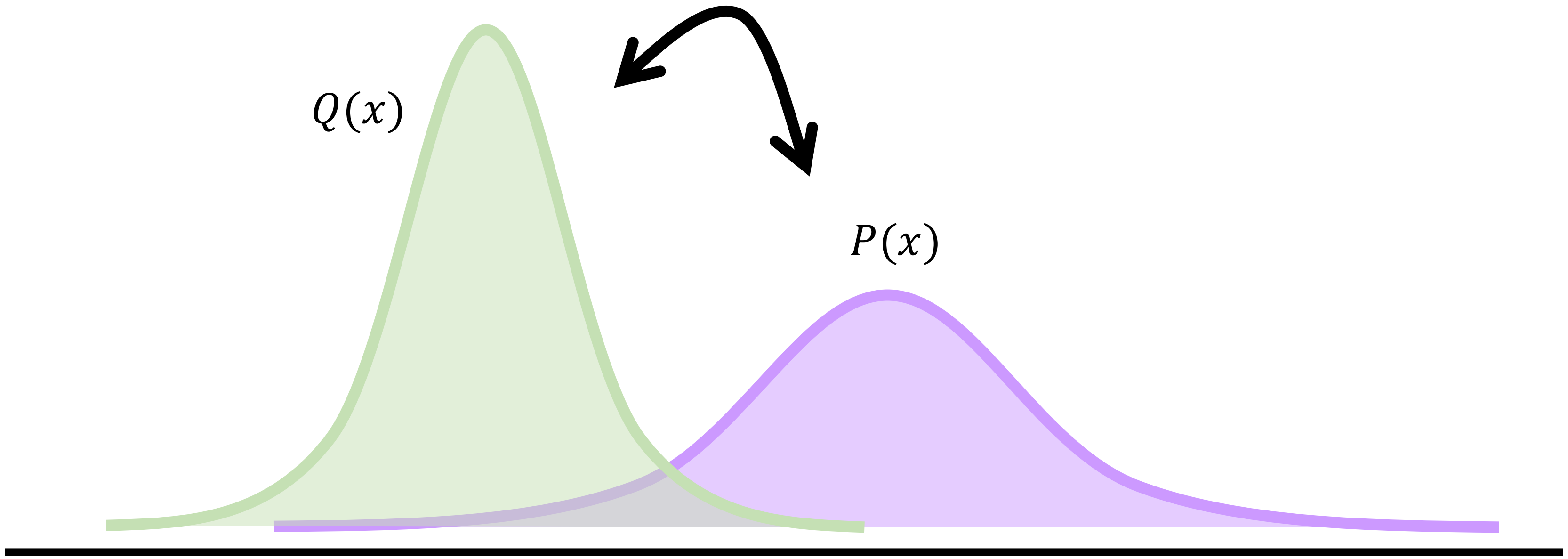
즉, 두 분포 사이의 차이에서 기준 분포 자체의 불확실성을 제거함으로써,



$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \underbrace{-\sum P(x) \log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x) \log P(x)\right)}_{\text{Entropy}}$$

두 분포가 얼마나 서로 다른 정보를 가지고 있는지를 순수하게 측정하는 방법입니다.



$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$$= \underbrace{-\sum P(x)\log Q(x)}_{\text{Cross Entropy}} - \underbrace{\left(-\sum P(x)\log P(x)\right)}_{\text{Entropy}}$$

그러면 이제는 KL Divergence를 계산하는 방법에 대해 간단한 예제를 통해 알아보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

신박AI

계산을 위해 다음과 같이 간단한 두 분포를 가정해보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

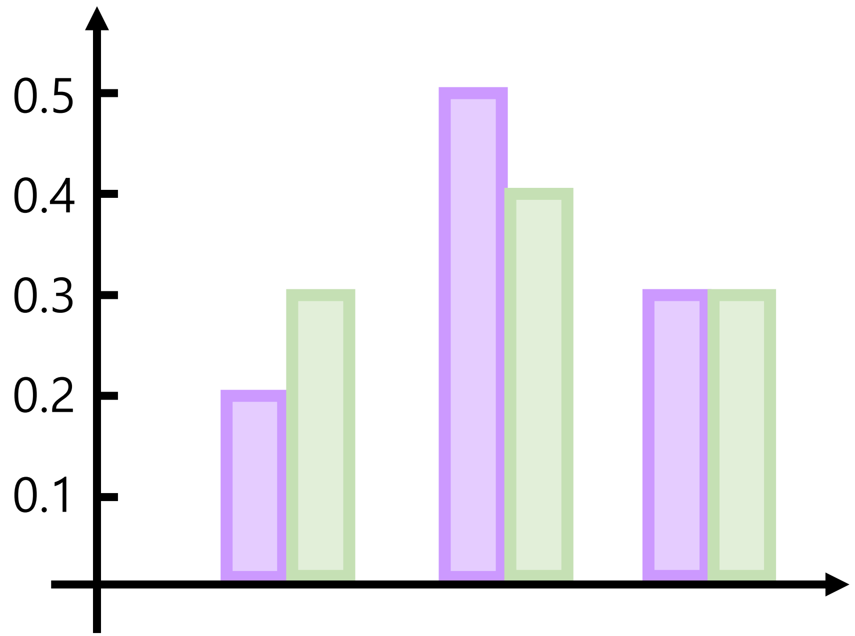
신박AI

계산을 위해 다음과 같이 간단한 두 분포를 가정해보도록 하겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

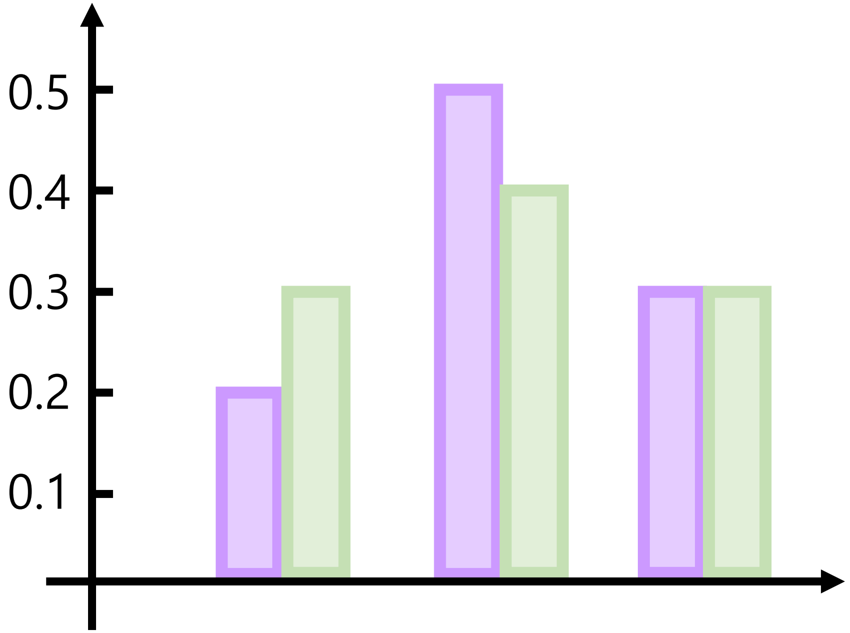
$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$

$Q(x) = \{\, 0.3, 0.4, 0.3 \,\}$



그러면 다음과 같이 식에 대입할 수 있고,

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

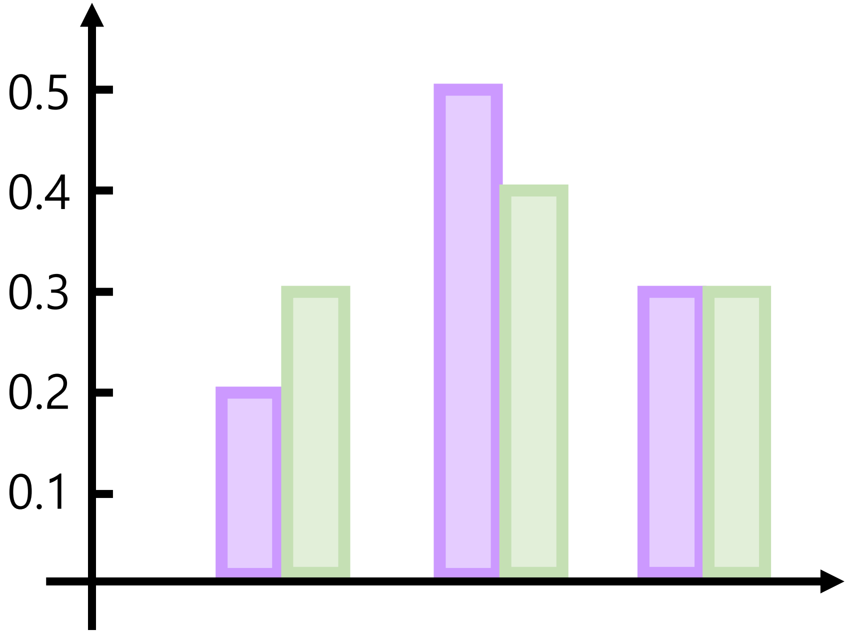


$P(x) = \{\ 0.2, 0.5, 0.3\ \}$ $Q(x) = \{\ 0.3, 0.4, 0.3\ \}$

$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.3}\right) + 0.5 \times \log\left(\frac{0.5}{0.4}\right) + 0.3 \times \log\left(\frac{0.3}{0.3}\right)$$

주어진 분포의 KL Divergence은 0.0305가 되겠습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$



$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$

$Q(x) = \{\, 0.3, 0.4, 0.3 \,\}$

$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.3}\right) + 0.5 \times \log\left(\frac{0.5}{0.4}\right) + 0.3 \times \log\left(\frac{0.3}{0.3}\right)$$

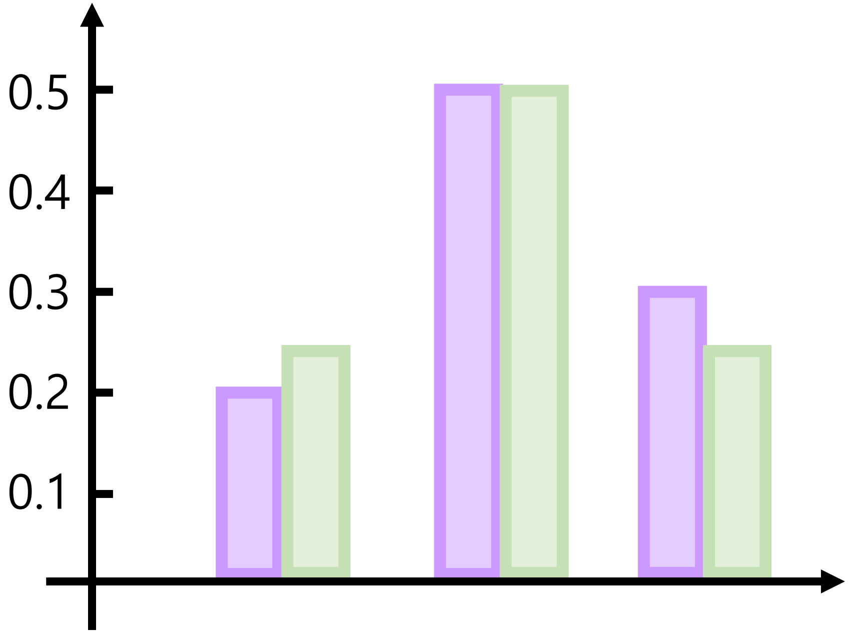
$$= 0.0305$$

신박AI

만약 두 분포가 더 서로 비슷해진다면 어떻게 될까요?

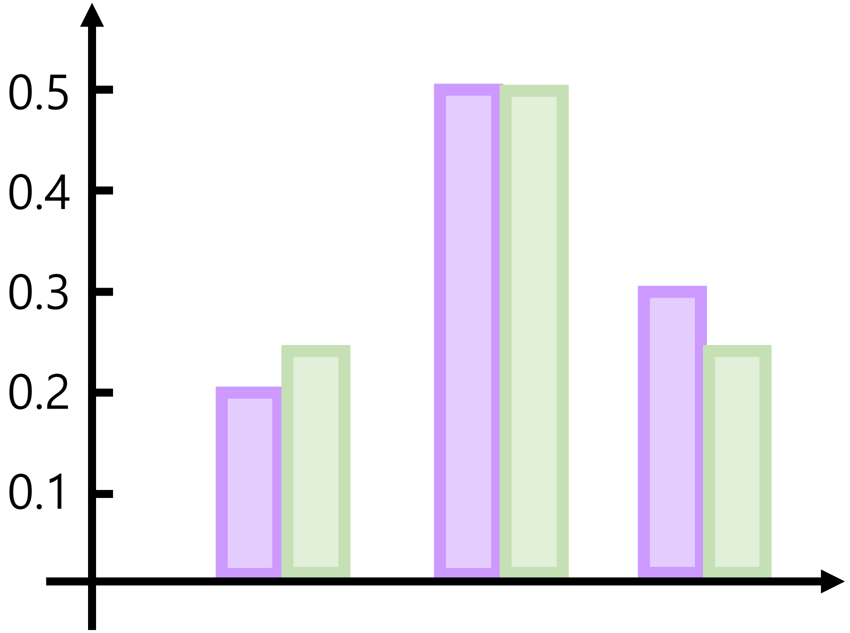
$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



그러면 다음과 같이 식에 대입할 수 있고,

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$



$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$

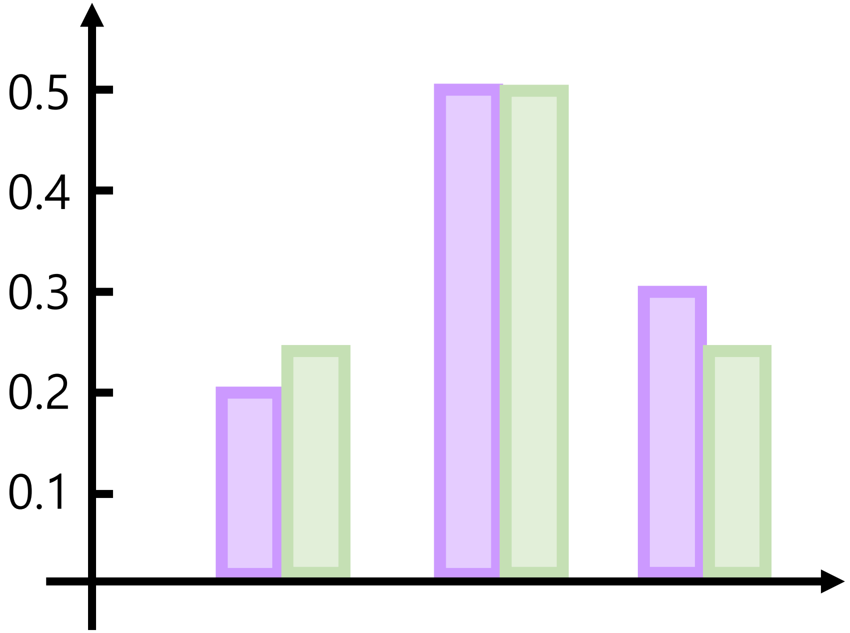
$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

KL Divergence값은 다음과 같이 0.01로 구할 수 있습니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$

$Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$

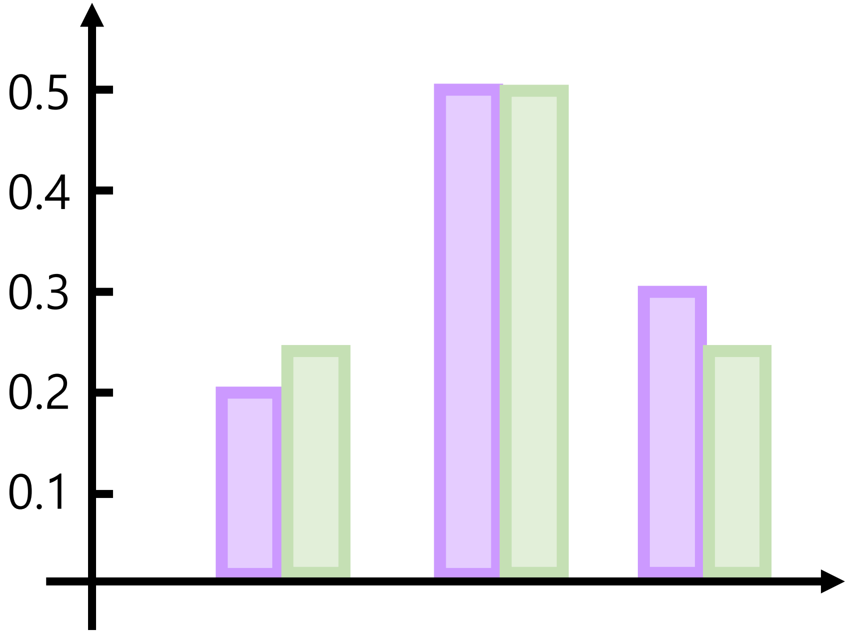


$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01$$

좀 전에 비해 KL Divergence값이 줄어든 것을 확인할 수 있습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$



$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$

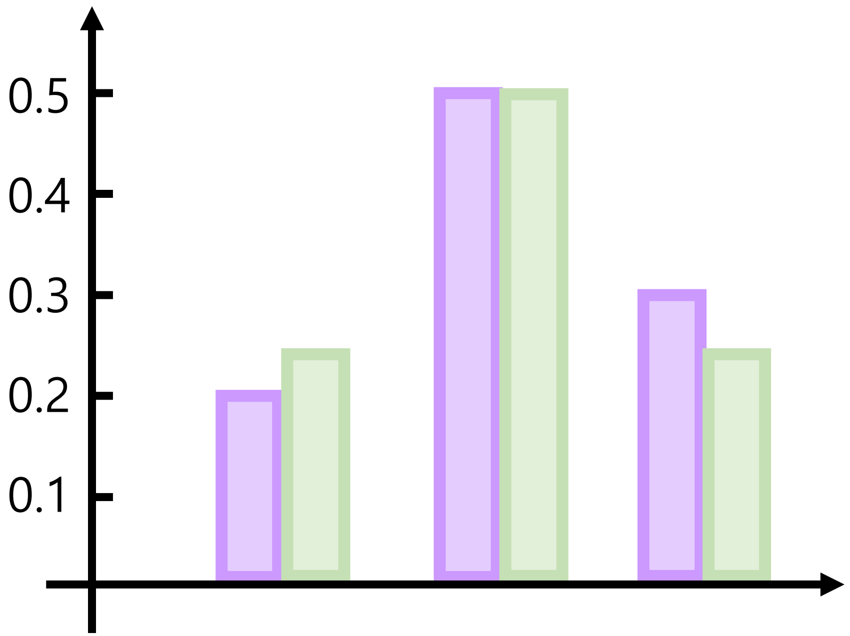
$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01 \quad \leftarrow 0.0305$$

신박AI

이와 같이 두 분포가 유사하면 KL Divergence의 값은 0에 수렴하게 됩니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$



$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$

$Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$

$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01 \quad \leftarrow 0.0305$$

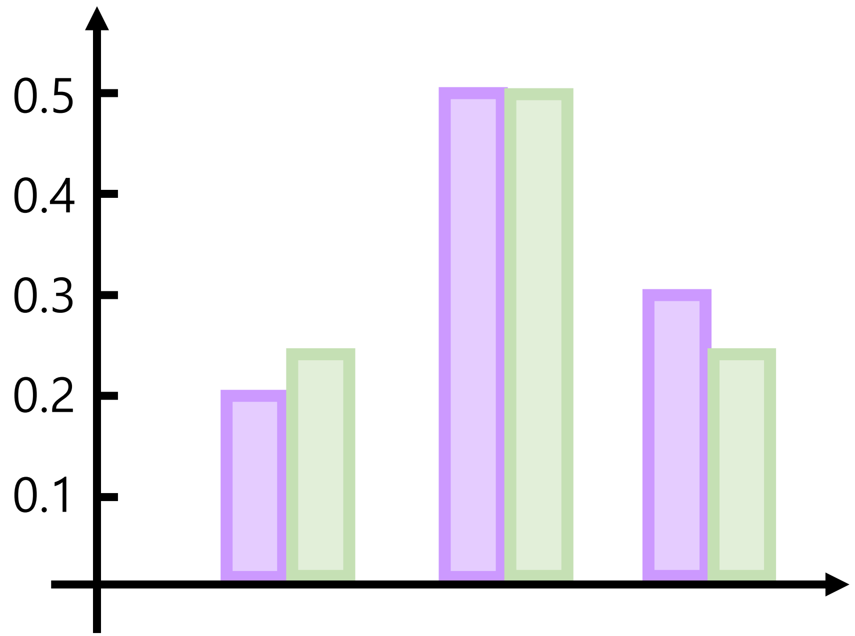
신박AI

뿐만 아니라, KL Divergence의 중요한 특징은 비대칭이라는 점입니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$

$Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



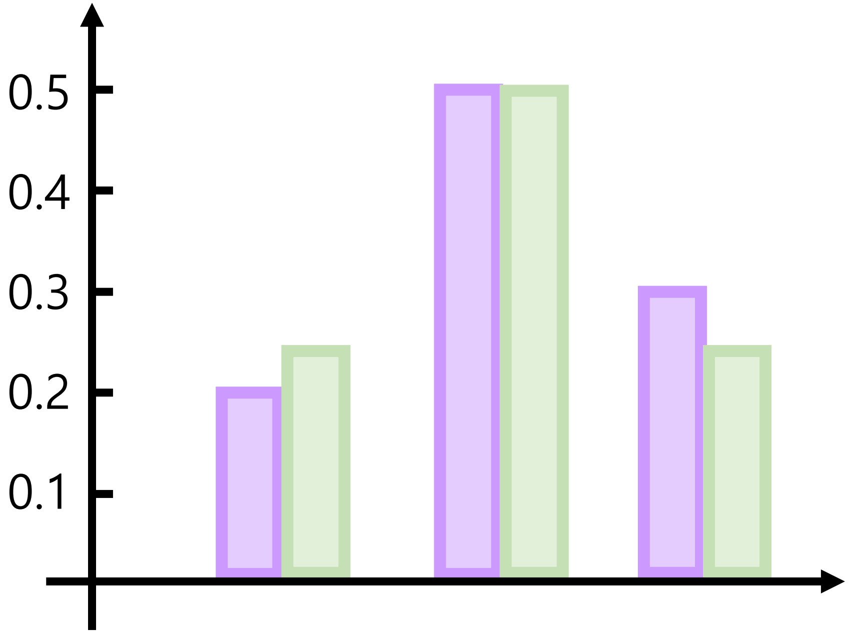
$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01$$

이와 같이 기준 분포가 P에서 Q로 바뀔 경우,

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \qquad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $\quad$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

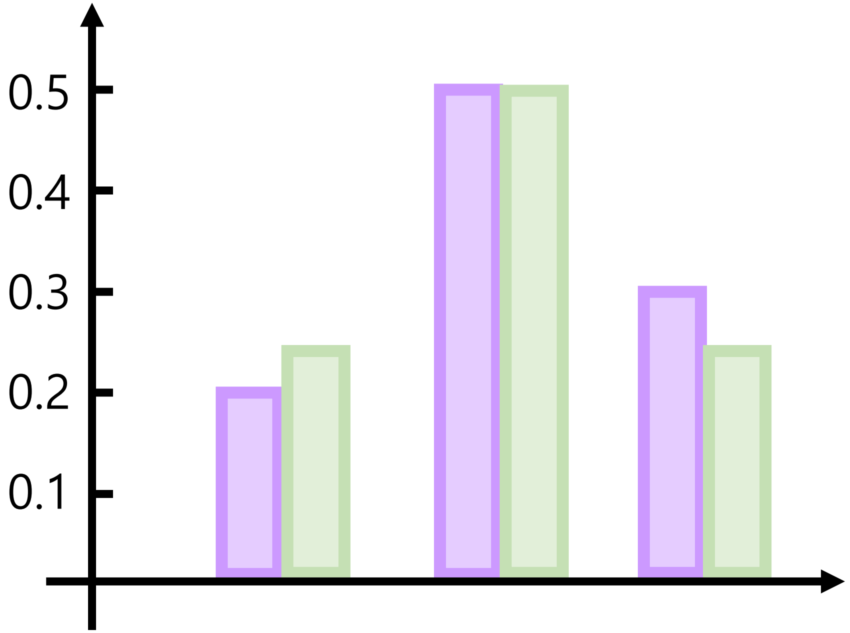
$$= 0.01$$

신박AI

같은 분포를 사용하여도 KL Divergence의 값은 달라집니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \qquad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $\quad$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01$$

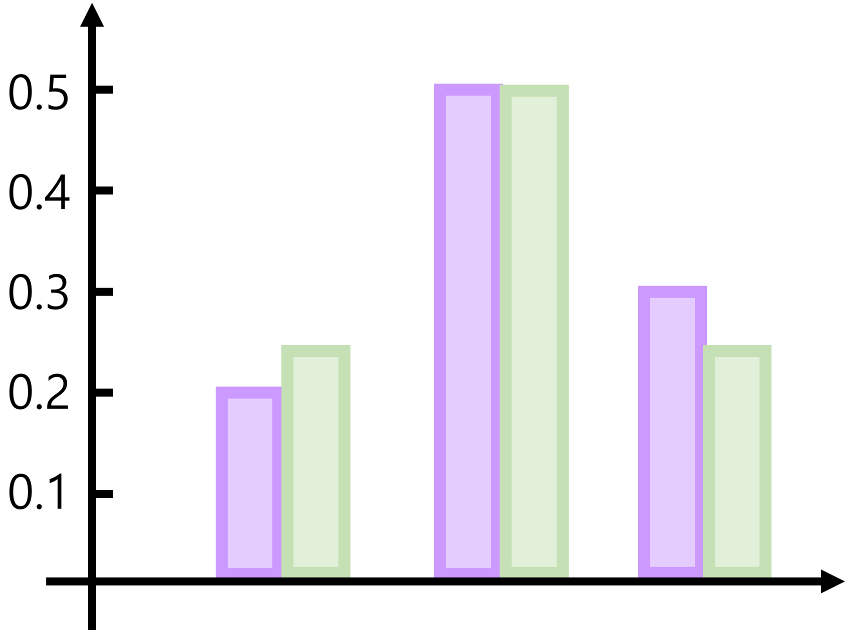
$$D_{KL}(Q \parallel P) = 0.25 \times \log\left(\frac{0.25}{0.2}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.25 \times \log\left(\frac{0.25}{0.3}\right)$$

$$= 0.0102$$

이는 KL Divergence가 두 분포의 불일치를 계산할 때,

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \qquad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$

$$= 0.01$$

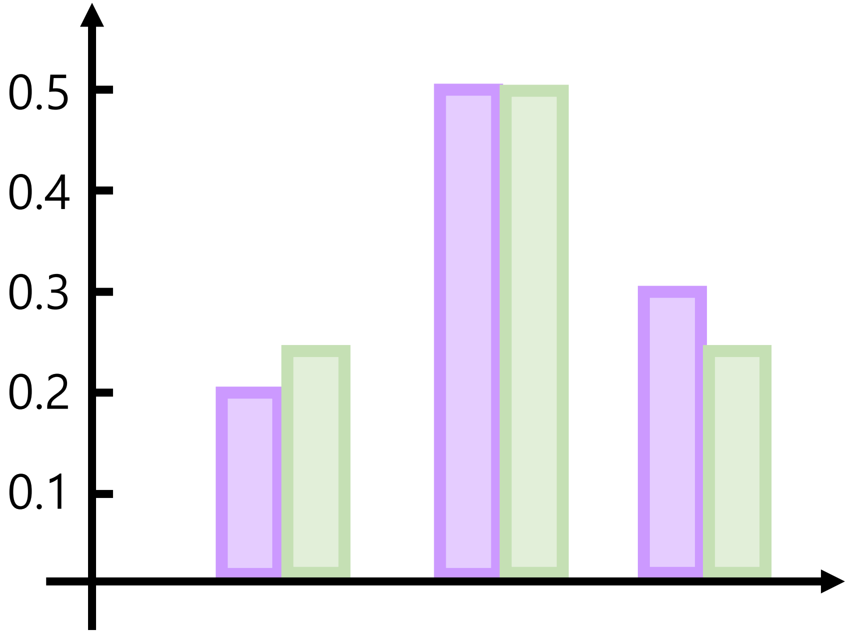
$$D_{KL}(Q \parallel P) = 0.25 \times \log\left(\frac{0.25}{0.2}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.25 \times \log\left(\frac{0.25}{0.3}\right)$$

$$= 0.0102$$

신박AI

특정 분포를 기준 분포로 하여 상대적인 차이를 계산하기 때문에 이런 특징을 보이게 됩니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right) \qquad D_{KL}(Q \parallel P) = \sum Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$



$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $\qquad$ $Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$

$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$
$$= 0.01$$

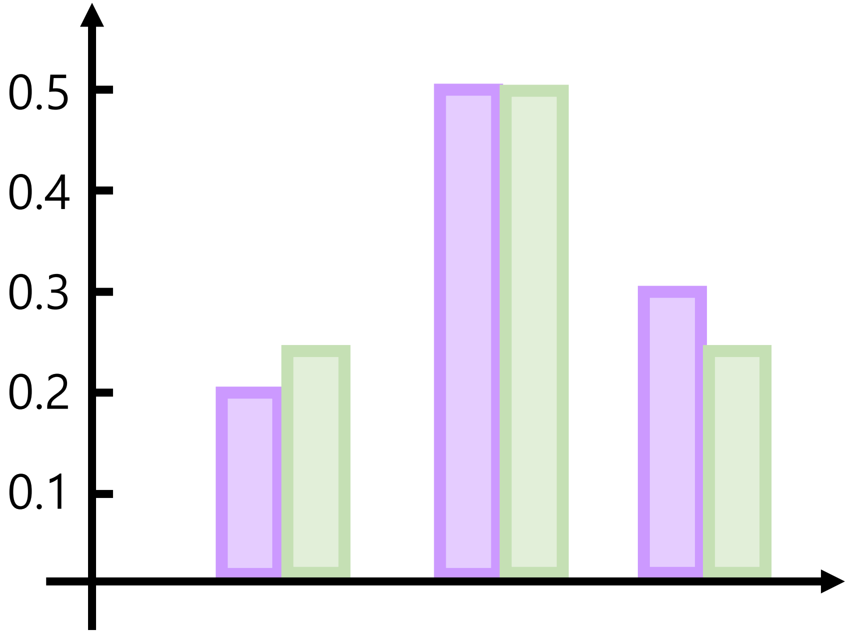
$$D_{KL}(Q \parallel P) = 0.25 \times \log\left(\frac{0.25}{0.2}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.25 \times \log\left(\frac{0.25}{0.3}\right)$$
$$= 0.0102$$

신박AI

이런 비대칭성은 때로는 데이터를 분석할 때, 다음과 같은 불편함이 있을 수 있습니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \qquad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$P(x) = \{\, 0.2, 0.5, 0.3 \,\}$ $\quad Q(x) = \{\, 0.25, 0.5, 0.25 \,\}$



$$D_{KL}(P \parallel Q) = 0.2 \times \log\left(\frac{0.2}{0.25}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.3 \times \log\left(\frac{0.3}{0.25}\right)$$
$$= 0.01$$

$$D_{KL}(Q \parallel P) = 0.25 \times \log\left(\frac{0.25}{0.2}\right) + 0.5 \times \log\left(\frac{0.5}{0.5}\right) + 0.25 \times \log\left(\frac{0.25}{0.3}\right)$$
$$= 0.0102$$

우선 KL Divergence는 두 분포의 차이를 계산할 때, 기준 분포를 고려해야 하고,

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \quad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

신박AI

또한 KL Divergence의 값의 범위는 0부터 무한대까지 나올 수 있으므로,

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right) \quad D_{KL}(Q \parallel P) = \sum Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$D_{KL}(P \parallel Q) = [0, \infty)$$

신박AI

분포의 차이를 계산하여, 모델을 업데이트 할 때 문제가 발생할 수도 있습니다.

$$D_{KL}(P \parallel Q) = \sum P(x) \log\left(\frac{P(x)}{Q(x)}\right) \quad D_{KL}(Q \parallel P) = \sum Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$D_{KL}(P \parallel Q) = [0, \infty)$$

신박AI

그래서 이를 보완한 척도로서 Jensen-Shannon Divergence가 있습니다.

$$D_{KL}(P \parallel Q) = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \quad D_{KL}(Q \parallel P) = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$$JSD = \frac{D_{KL}(P \parallel Q) + D_{KL}(Q \parallel P)}{2}$$

신박AI

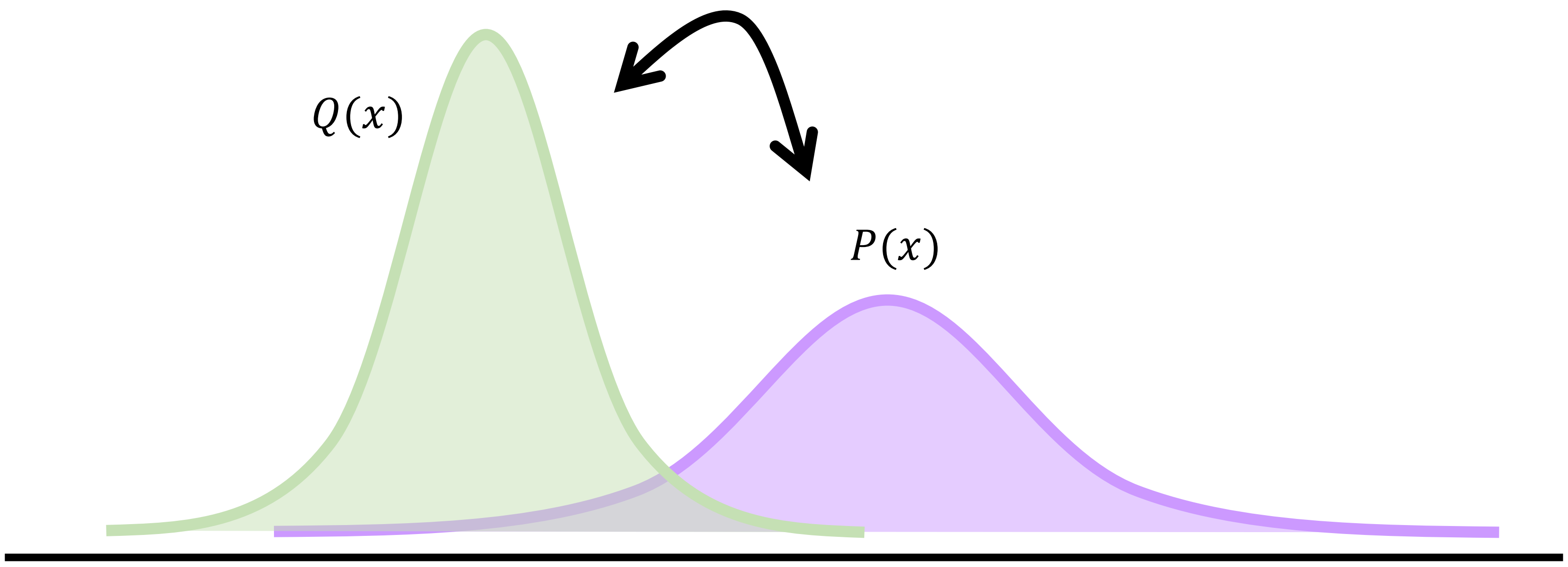
Jensen-Shannon Divergence는 두 분포가 만드는 KL Divergence 값들을 서로 합하여 2로 나눈 값으로

$$\boxed{D_{KL}(P \parallel Q)} = \sum P(x)\log\left(\frac{P(x)}{Q(x)}\right) \qquad \boxed{D_{KL}(Q \parallel P)} = \sum Q(x)\log\left(\frac{Q(x)}{P(x)}\right)$$

$$JSD = \frac{\boxed{D_{KL}(P \parallel Q)} + \boxed{D_{KL}(Q \parallel P)}}{2}$$

신박AI

KL Divergence에 비하여 훨씬 두 확률분포 사이의 상대적인 차이를 부드럽게 측정 할 수가 있습니다.



$$JSD = \frac{\boxed{D_{KL}(P \parallel Q)} + \boxed{D_{KL}(Q \parallel P)}}{2}$$

신박AI

오늘 제가 준비한 KL Divergence에 관한
영상은 여기까지입니다.

신박AI

KL Divergence의 개념을 직관적으로 또
쉽게 이해하는데 중점을 두어
짧은 영상을 준비해보았습니다.

영상 끝까지 시청해 주셔서 감사드립니다.
다음에도 흥미롭고 도움이 되는 주제로
찾아뵙겠습니다.

딥러닝과 머신러닝을 공부하시는데
도움이 되셨기를 희망합니다!
다음 시간에 또 만나요!

신박AI

감사합니다!

좋은 하루 되세요!!

이 채널은 여러분의 관심과 사랑이 필요합니다

신박AI

‘좋아요’와 ‘구독’버튼은 강의 준비에 큰 힘이 됩니다!

그리고 영상 자료를 사용하실때는
출처 ‘신박AI’를 밝혀주세요

신박AI