KL-Divergence

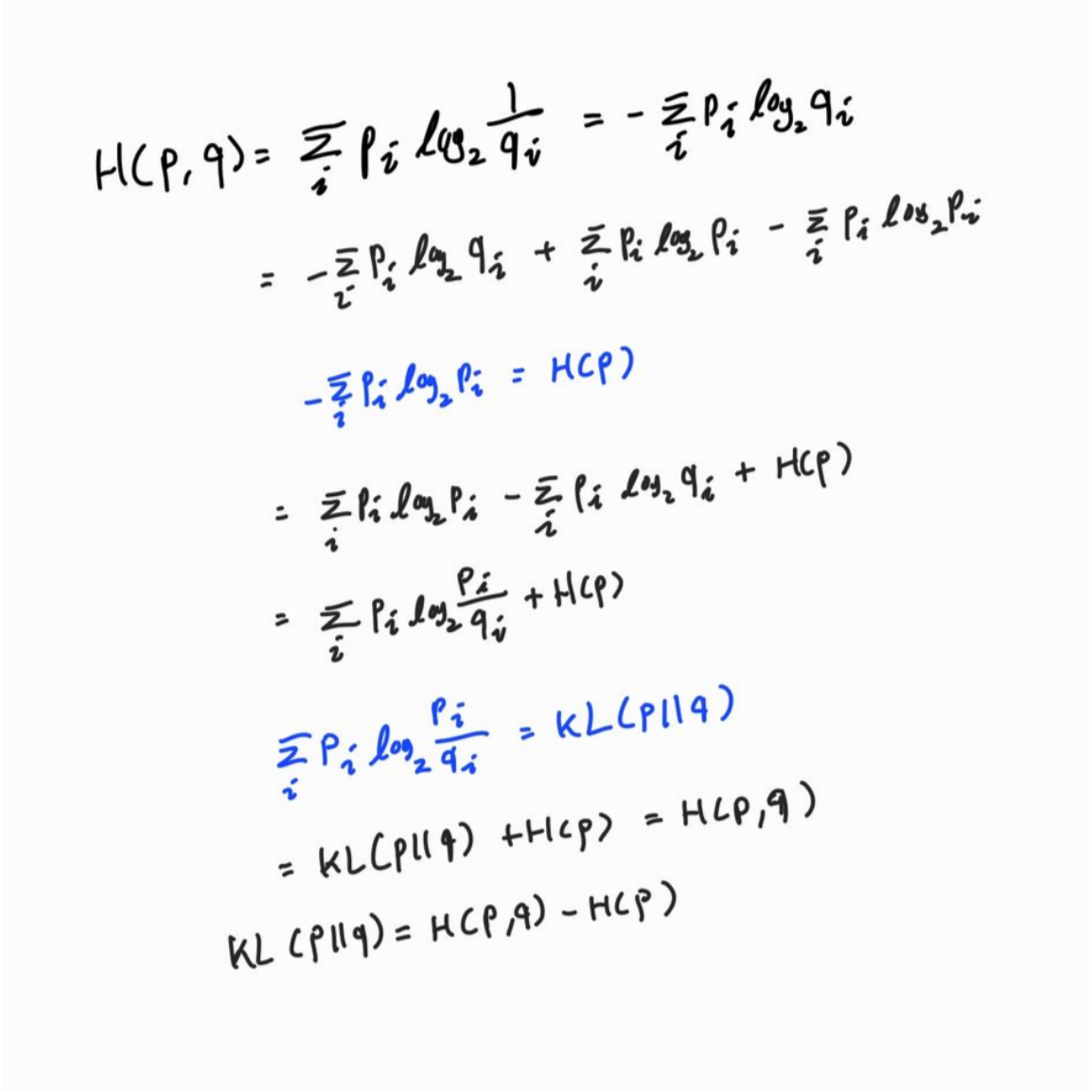
크로스 엔트로피 $H(p,q)$는 모델에서 예측한 확률(p)과 정답확률(q)을 모두 사용해 측정한 값이다.

학습이 진행될 수록 $H(p,q)$는 줄어든다.

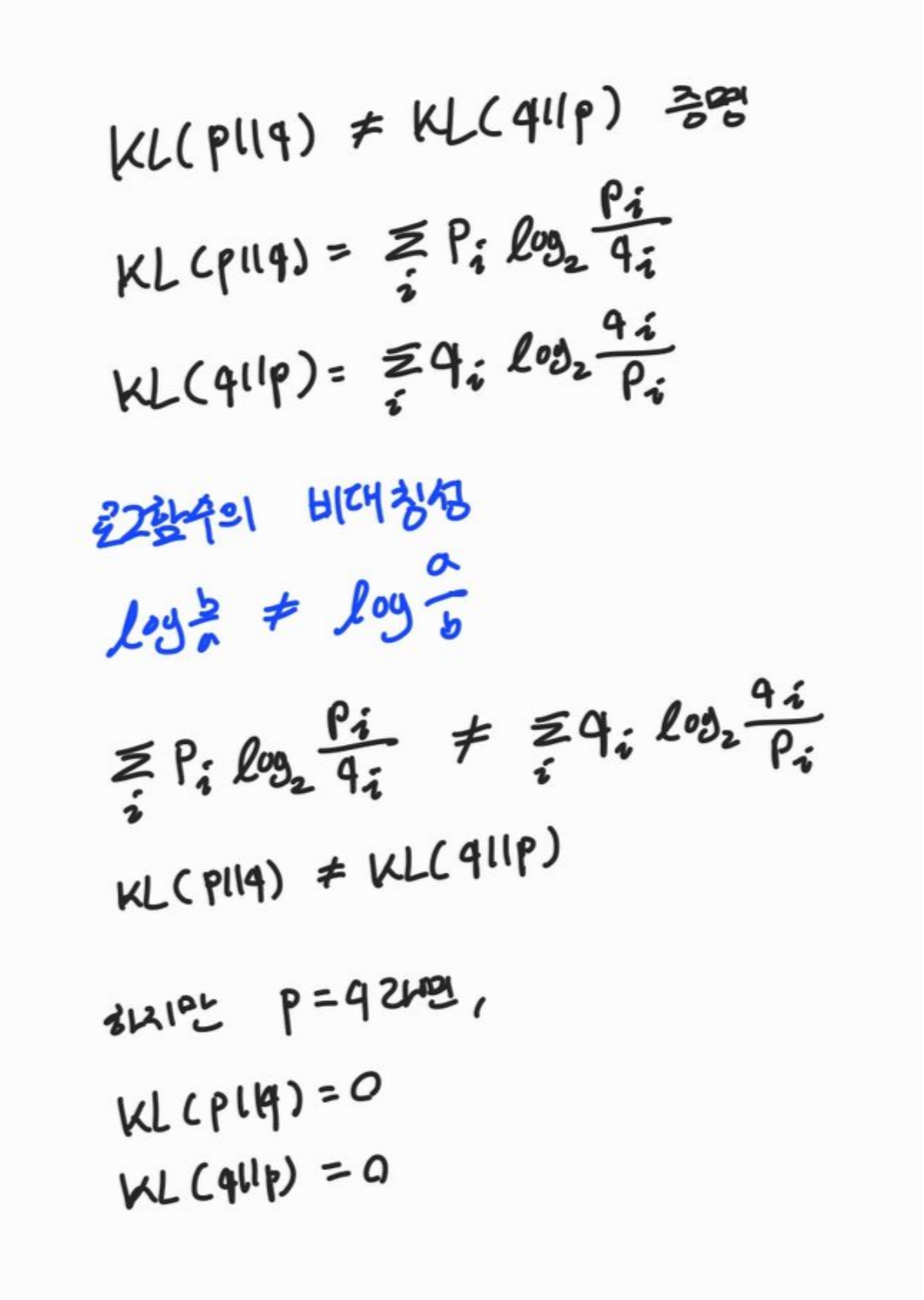
KL-Divergence는 두 확률분포의 차이를 계산한다. 어떤 이상적인 분포에 대해, 그 분포를 근사하는 다른 분포를 사용해 샘플링을 한다면 발생할 수 있는 엔트로피 차이이다.

다음 3가지 특성을 가지고 있다.

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



2. $KL(p||q) \geq 0$



3. $KL(p||q) \neq KL(q||p)$

$KL(p||q) \geq 0$ 증명

$$KL(p||q) = -\sum_i p_i \log_2 \frac{q_i}{p_i}$$

Convex 함수 $f(x) = -\log x$

젠슨 부등식 $E[f(x)] \geq f(E[x])$

$$E[f(x)] = -\sum_i p_i \log_2 \frac{q_i}{p_i}$$

$$f(E[x]) = -\log_2 \left(\sum_i p_i \frac{q_i}{p_i}\right)$$

$$= -\log_2 \left(\sum_i q_i\right)$$

$$= -\log_2 (1) = 0$$

$E[f(x)] \geq 0$

$KL(p||q) \geq 0$