변분법적 추론(variational inference)

<상황>

: likelihood와 prior을 아는 상황일 때, Posterior을 구하고 싶음(MAP를 구하는 것은 가능)

$$posterior(p(w|D)) = \frac{p(D|w)p(w)}{\int p(D|w)p(w)dw} = \frac{likelihood*prior}{evidence}$$

-> likelihood와 prior을 알더라도 수많은 weight에 대한 적분을 하는 것은 불가능에 가깝기 때문에 evidence를 구하기 힘듦(NN의 layer가 많아질수록 weight는 더욱 많아짐)

<Variational inference>

모델 f에 대한 $p(w|D)$를 구하기는 힘들기 때문에, 우리가 잘 알고 비교적 미지수가 적은 함수 g $q(w|\theta)$를 통해 $p(w|D)$를 추론

-> $p(w|D)$와 $g(w|\theta)$가 최대한 비슷한 분포이어야 하기 때문에, 두 확률분포의 유사도에 대한 척도로 KL-Divergence사용. 이 척도를 **최소화** 하는 것이 목표

$$D_{KL}[q(w|\theta)\|p(w|D)] = \int q(w|\theta)\log\frac{q(w|\theta)}{p(w|D)}dw$$

$$\int q(w|\theta)\log\frac{q(w|\theta)}{p(w|D)}dw$$
$$= \int q(w|\theta)\log\frac{q(w|\theta)p(D)}{p(w)p(D|w)}dw$$
$$= \int q(w|\theta)[\log q(w|\theta)p(w) + \log p(D) - \log p(D|w)]dw$$
$$= \int q(w|\theta)\log\frac{q(w|\theta)}{p(w)}dw + E_q[\log p(D)] - E_q[\log p(D|w)]$$
$$= D_{KL}[q(w|\theta)\|p(w)] + E_q[\log p(D)] - E_q[\log p(D|w)]$$

이 때, $Eq[\log p(D)]$의 경우 상수이기 때문에 무관.

<Maximize ELBO>

$$D_{KL}[q(w|\theta)\|p(w|D)] = D_{KL}[q(w|\theta)\|p(w)] + E_q[\log p(D)] - E_q[\log p(D|w)]$$
$$\log p(D) = D_{KL}[q(w|\theta)\|p(w|D)] + E_q[\log p(D|w)] - D_{KL}[q(w|\theta)\|p(w)]$$
$$\log p(D) \geqq E_q[\log p(D|w)] - D_{KL}[q(w|\theta)\|p(w)]\ (\because D_{KL}[q(w|\theta)\|p(w|D)] \geqq 0)$$

- Evidence Lower Bound(ELBO) : $E_q[\log p(D|w)] - D_{KL}[q(w|\theta)\|p(w)]$

-> $KLD = evidence - ELBO$ -> ELBO를 최대화