변분법적 추론

잠재변수 모형

우리가 원하는 것은 확률적 데이터 X에 대한 확률모형 즉 확률분포 $p(X)$를 찾는 것이다. 변분법적 추론에서는 두 가지 가정을 한다.

- X는 잠재변수 Z의 영향을 받는 네트워크 모형으로 $p(X|Z)$는 가정에 의해 주어져 있다. 따라서 잠재변수 확률분포 $p(Z)$를 구하면 $p(X)$도 구할 수 있다.
- 확률분포 $p(Z)$를 직접 구하기 어려우므로 유사한 확률분포 $q(Z)$를 찾는다.

우선 다음 공식을 증명할 수 있다. 이 식은 EM 알고리즘에서 설명한 것과 유사하지만 잠재변수 Z가 모수 θ를 포함하고 연속확률변수인 경우를 감안하여 합이 아닌 적분을 사용하였다.

$$\log p(X) = \int q(Z) \log \left(\frac{p(X,Z)}{q(Z)} \right) dZ - \int q(Z) \log \left(\frac{p(Z|X)}{q(Z)} \right) dZ$$

증명은 EM 방법의 경우와 같다.

이제부터 첫 항은 $L(q)$, 두번째 항은 $KL(q\|p)$라고 쓰도록 한다.

$$L(q) = \int q(Z) \log \left(\frac{p(X,Z)}{q(Z)} \right) dZ$$

$$KL(q\|p) = - \int q(Z) \log \left(\frac{p(Z|X)}{q(Z)} \right) dZ$$

$L(q)$는 분포함수 $q(Z)$를 입력하면 수치가 출력되는 범함수(functional)이다. $KL(q\|p)$은 분포함수 $q(Z)$와 $p(Z|X,\theta)$의 차이를 나타내는 쿨백-라이블러 발산이다. 쿨백-라이블러 발산은 항상 0과 같거나 크기 때문에 $L(q)$는 $\log p(X)$의 하한(lower bound)가 된다.

$$\log p(X) \geq L(q)$$

반대로 이야기하면 $\log p(X)$가 $L(q)$의 상한이다.

그리고 이 때 쿨백-라이블러 발산은 0이 된다. 따라서 L을 최대화(쿨백-라이블러 발산을 최소화)하는 분포함수 q를 찾아낼 수 있다면 가능도 $\log p(X)$를 최대화하는 것과 같다.