15 강 확률분포(3)

◈ 담당교수: 장필훈

▣ 주요용어

용어	해설
감마분포	확률밀도함수가 $\frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}, x > 0$로 주어지는 분포.
정규감마분포	정규분포와 감마분포의 곱.
스튜던트분포	정규감마분포에서 정밀도에 대해 marginalize 한 분포. 모든 분산에 대한 가우시안의 총 합을 나타낸다.
켤레사전분포	사전확률과 사후확률이 같은 분포계열에 속할 때 그 사전확률분포를 가리키는 말. 계산상 잇점이 있다.(가능도함수와 곱을 매번 계산하지 않고 매개변수만 업데이트 하는 방식으로 사후확률을 계산해낼수있다)

▣ 정리하기

1. (가우시안의) 평균을 아는 상태에서 분산을 추정할 때, 켤레 사전분포는 감마분포이다.
    a. 사전분포와 사후분포를 비교해보면 (같은 감마분포이고) 다음과 같은 변화를 관찰 할 수 있다.

       $a_N = a_0 + \frac{N}{2},\ \ b_N = b_0 + \frac{N}{2}\sigma_{ML}^2$
    b. 사전분포의 매개변수 a_0는 $2a_0$개의 사전관측값이라고 해석할 수 있다.

2. 평균과 분산을 모두 모르는 상태라면 켤레사전분포는 다음의 정규감마(가우시안 감마)이다.

$$\mathcal{N}\left(\mu \middle| \mu_0, \frac{1}{\beta\lambda}\right) \mathrm{Gam}(\lambda|a, b)$$

3. 다변량 가우시안의 경우
    a. 분산을 알 때 평균에 대한 켤레사전분포는 가우시안이다.
        i. 단변량일때와 같다.
    b. 평균을 알 때 공분산행렬에 대한 켤레사전분포는 ‘위샤트 분포’라고 한다.

4. 같은 평균과 다른 정밀도를 가진 무한히 많은 가우시안 분포를 합산하면 스튜던트분포(t 분포)가 된다.
    a. 가우시안을 무한히 혼합한 것이기 때문에 outlier 에 대해 robust 하게 된다.
    b. 최대가능도해는 EM 을 이용해 구한다.

5. 극좌표계를 이용해서 주기적 변수로 확률변수로 나타낼 수 있다.

6. 실제 데이터집합은 복잡하기 때문에 단일모델로 나타내기 어려운 경우가 많고, 그럴때는 혼합모델을 사용한다.

7. $h(x)g(\eta)\exp\left(\eta^T u(x)\right)$ 로 나타낼 수 있는 모든 분포를 지수족 분포라고 한다.
    a. 일례로, 베르누이 분포를 지수족분포의 형태로 나타낼 수 있다.
    b. 다항분포도 지수족 분포의 형태로 나타낼 수 있다.
    c. 가우시안분포도 지수족이다.

8. 켤레사전분포를 이용할 수 있으면, 사후확률을 계산할 때 파라미터만 업데이트 하는 방법으로 계산할 수 있다.

9. 모든 지수족 분포에 적용 가능한 켤레사전분포를 정의할 수 있다.

10. 해석적 분포를 사용할 수 없으면, 파라미터 없이 밀도추정을 해야 한다.
    a. 예: 히스토그램 밀도추정
        i. 구간너비에 따라 결과가 달라진다.
        ii. 고차원 데이터를 다루기 부적합하다.
    b. 크게, 커널밀도추정과 K 최근접 이웃 두가지로 구분해볼 수 있다.