Ch 02. 최대 가능도 추정

1. 생성 모델 개요
2. 실제 데이터로 생성 모델 구현하기
3. 최대 가능도 추정 이론
4. 생성 모델의 용도

2.1 생성 모델 개요

2.1.1 생성 모델의 목표

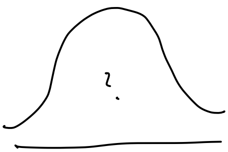
$p(x; ?, ?)$

샘플링 → $(x^{(1)}, x^{(2)}, \cdots, x^{(N)})$

← 어디서 왔지?

↳ 이걸 찾는 게 생성 모델의 목표! ⇒ 이러한 과정을 추론(inference) 혹은 매개변수 추정(parameter estimation)

2.1.2 모집단과 샘플



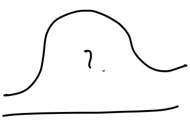
모집단 분포! (population distribution)
전체 집합!

$P_*(x)$ ← 너무 알고 싶어 ㅠㅠ 그럼, 아래!! 처럼!

모델링! → 1. $P_*(x)$가 뭔지 몰라. 그러면 대충 매개변수로 표현되는 분포라고 가정 하자!

매개변수 추정! → 2. 그리고 $P_*(x)$에서 얻은 샘플들이 $P_*(x)$에 맞게 매개변수를 조정하자!

2.3 최대 가능도 추정 이론



샘플링 → $\{x^{(1)}, x^{(2)}, \cdots, x^{(N)}\}$

이때!

1. 모델링부터 → 정규 분포라고 하자! → 그럼 매개변수는 μ, σ!

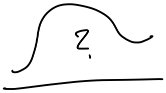
2. 매개변수 추정을 하자 → 그러하면!

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x^{(n)}$$

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (x^{(n)} - \hat{\mu})^2}$$

왜인지는 뒤에서! ←

2.3.1 가능도 최대화



→ 이때! 매개변수 = θ 라면, 데이터 x을 얻을 확률밀도 = $p(x;\theta)$

그럼!

$\xrightarrow{\text{샘플링}}$ $D = \{x^{(1)}, x^{(2)}, \cdots, x^{(N)}\}$ 일 때!

이걸 얻을 확률 밀도 $= p(D;\theta) = p(x^{(1)};\theta)\, p(x^{(2)};\theta) \cdots p(x^{(N)};\theta)$

$$= \prod_{n=1}^{N} p(x^{(n)};\theta)$$

매개변수 = θ 일 때, D를 얻을 확률이닷!

이걸, θ를 인수로 받는 함수!라고 하면! 이렇게 표현!

$L(\theta) = p(D;\theta)$ = 가능도 (likelihood) ★중요★

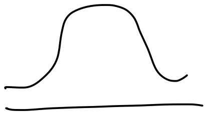
= 가능도 함수(likelihood function)

여기서*

$\log L(\theta) = \log p(D;\theta)$ = 로그 가능도★

자몽 여기서 가능도의 최대화란?
θ를 잘 찾아서?
D가 딱! 맞는 분포를 찾는거지?
즉! ★ D를 얻을 확률을 ★
'최대화' 하는 것이로다! ★
★

2.3.2 정규 분포의 최대 가능도 추정



$$\mathcal{N}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\xrightarrow{\text{샘플링}} D = \{x^{(1)}, x^{(2)}, \cdots, x^{(N)}\} \xrightarrow{\text{확률밀도}} \underbrace{p(D;\mu,\sigma) = \prod_{n=1}^{N} p(x^{(n)};\mu,\sigma)}_{\text{이게 가능도!}}$$

$$\log p(D;\mu,\sigma) = \log \prod_{n=1}^{N} p(x^{(n)};\mu,\sigma)$$

$$= \underbrace{-\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x^{(n)}-\mu)^2}_{\text{이게 로그 가능도!}}$$

자! 여기서! 최대 가능도를 어떻게 얻을까!?

→ 어쨌든 $\log p(D;\mu,\sigma)$가 최대화 되는 μ, σ를 찾아야 해

→ 2차 미분?

→ 2차 미분값 = 0 이 되는 곳?

→ 아... 2차 조건이... 2차 함수.. 2차 계수가 ⊖인...

→ 오? 로그 가능도를 보니?....!!!

→ 찾았드아!!!

결국: μ와 σ로 편미분 하면!!! → 아래식! → 오! 증명됐드아!!!

$$\frac{\partial \log p(D;\mu,\sigma)}{\partial \mu} = 0$$

$$\Leftrightarrow \mu = \frac{1}{N}\sum_{n=1}^{N} x^{(n)}$$

$$= \hat{\mu}$$

$$\frac{\partial \log p(D;\mu,\sigma)}{\partial \sigma} = \frac{\partial \log p(D;\mu=\hat{\mu},\sigma)}{\partial \sigma}$$

$$\Leftrightarrow \sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x^{(n)}-\hat{\mu})^2}$$

$$= \hat{\sigma}$$