

★ $\frac{1}{2} \frac{1}{4} \frac{1}{4}$, 따라가기 $\frac{1}{2}$, 시간 다.

• 유이득증상: 유방 크기가 커짐, 사슴의 유방이 커짐. 유방의 크기가 커짐. 유방의 크기가 커짐.

MLE $\rightarrow$ MAP $\rightarrow$ Bayesian posterior predictive $\rightarrow$ ~~Bayesian linear regression.~~

Inclass 19: Posterior Predictive Distribution

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

Department of Artificial Intelligence, Dongguk University

ଅ(କ୍ଷ):

① coss toin.
(discrete)
 $\theta_1, \theta_2.$

② Gaussian.
(continuous)
python.

등전면하기.

$\checkmark H$	$\checkmark T$
θ	$1-\theta$

parameter θ 를 이용하여
확률분포를 표현함.

문제: 주어진 data $\rightarrow \theta$ 를 추정해보자.

parameter를

특정값으로

결정한다.

① likelihood $\rightarrow$ MLE

② posterior $\rightarrow$ MAP estimate.
(prior, likelihood)

③ posterior predictive.

parameter를 여러

값을 갖지 않는다.

두개의 동전 확률: $\theta_1 = 0.5$

$\theta_2 = 0.9$

관측값 5개 :

H T T T T

① likelihood, MLE.

$$\mathcal{L}(\theta_1) = p(\underline{\text{HTTTT}}; \theta_1) = p(H; \theta_1) p(T; \theta_1)^4 = \underline{0.00125}$$

$$\mathcal{L}(\theta_2) = p(\text{HTTTT}; \theta_2) = p(H; \theta_2) p(T; \theta_2)^4 = \underline{0.00009}$$

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \{\theta_1, \theta_2\}} \mathcal{L}(\theta) = \theta_1 = \underline{0.5}$$

$$p(\underline{H}) = p(H; \hat{\theta}_{\text{MLE}}) = \theta_1 = \underline{0.5}$$

$$p(T) = p(T; \hat{\theta}_{\text{MLE}}) = 1 - \theta_1 = \underline{0.5}$$

② posterior, MAP.

$\left[\bigoplus \text{가정 추가. } \theta \in \text{최소 2개 가능} \right]$

$$P(B|A) \propto P(A|B)P(B) \\ = \frac{P(A|B)P(B)}{\sum_B P(A|B)P(B)}$$

[prior distribution $p(\theta_1) = 0.05$ $p(\theta_2) = 0.95$]

$$p(\theta_1 | HTTTT) = \frac{p(HTTTT | \theta_1) p(\theta_1)}{p(HTTTT | \theta_1) p(\theta_1) + p(HTTTT | \theta_2) p(\theta_2)} = 0.422$$

+

$$p(\theta_2 | HTTTT) = \frac{p(HTTTT | \theta_2) p(\theta_2)}{p(HTTTT | \theta_1) p(\theta_1) + p(HTTTT | \theta_2) p(\theta_2)} = 0.578$$

11

1

$$P(\theta_1 | HTTTT) = 0.422$$

$$P(\theta_2 | HTTTT) = 0.578$$

✓

HTTTT

	H	T
θ_1	0.5	0.5
θ_2	0.9	0.1

$$\hat{\theta}_{MAP} = \arg \max_{\theta \in \{\theta_1, \theta_2\}} P(\theta | HTTTT) = \theta_2 = 0.9$$

$$P(H; \hat{\theta}_{MAP}) = \theta_2 = 0.9$$

$$P(T; \hat{\theta}_{MAP}) = 1 - \theta_2 = 0.1$$

$$\begin{array}{ccc}
 \begin{array}{l}
 \text{0.422} \\
 \uparrow \\
 \frac{P(\theta_1 | H T T T T)}{P(\theta_2 | H T T T T)} \\
 \text{0.578}
 \end{array}
 &
 \begin{array}{l}
 \text{1.25e-3} \\
 \frac{P(H T T T T | \theta_1)}{P(H T T T T | \theta_2)} \\
 \text{9e-5}
 \end{array}
 &
 \begin{array}{l}
 \frac{P(\theta_1)}{P(\theta_2)} \\
 \text{0.30} \\
 \text{0.70}
 \end{array}
 \end{array}$$

posterior = likelihood x prior

우리가 원하는 것의 곱셈 계산.

③ posterior predictive distribution.

가능한 모든 경우를
고려해볼 때의 확률

$$p(\theta_1 | H T T T T) = 0.422 \quad \text{posterior.}$$

$$p(\theta_2 | H T T T T) = 0.578$$

θ_1 일 때 H일 확률 = 0.5

꼭지진관측하거

θ_1 일 때 H일 확률. 0.422.

$$p(X=H | \theta_1) p(\theta_1 | H T T T T)$$

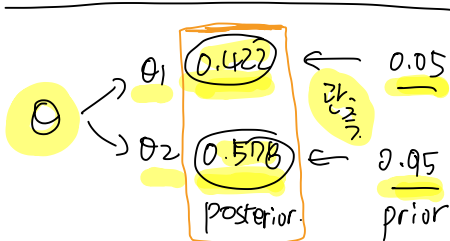
$$p(X=H) =$$

$$+ p(X=H | \theta_2) p(\theta_2 | H T T T T)$$

θ_2 일 때 H일 확률. 0.9

꼭지진관측하거

θ_2 일 때 H일 확률.



$$= 0.211 + 0.520$$

$$= 0.731$$

① MLE

$$\hat{\theta}_{MLE} = \theta_1 = 0.5$$

$$P(H) = 0.5$$

② MAP

$$\hat{\theta}_{MAP} = \theta_2 = 0.9$$

$$P(H) = 0.9$$

③ posterior predictive

$$P(\theta_1 | \text{data}) = 0.422$$

$$P(\theta_2 | \text{data}) = 0.576$$

$$P(H) = 0.731$$

$$\begin{aligned} &= P(H|\theta_1) P(\theta_1 | \text{data}) + P(H|\theta_2) P(\theta_2 | \text{data}) \\ &\quad \begin{array}{cc} \swarrow \text{data} & \searrow \text{data} \\ \uparrow \text{data} & \uparrow \text{data} \end{array} \end{aligned}$$

MLE, MAP

주어진 관측 $X_1, X_2, \dots, X_N$

$\downarrow$

$\frac{p(X_1, X_2, \dots, X_N | \theta)}{p(\theta | X_1, \dots, X_N)}$

파라미터 추정 ← 관측의 역할 끝

한 관측의 확률

$p(X | \hat{\theta})$

posterior predictive.

한 관측의 확률

$p(X | X_1, X_2, \dots, X_N)$

$= \int p(X | \theta) p(\theta | X_1, \dots, X_N) d\theta$

posterior

가능한 모든 경우의
이 확률을 계산하라. 0/1

posterior predictive.

θ data.

$$\Rightarrow \underline{p(x | x_1, x_2, \dots, x_N)} = \int \underbrace{p(x | \theta)}_{\text{model.}} \underbrace{p(\theta | x_1, x_2, \dots, x_N)}_{\text{likelihood} \times \text{prior.}} d\theta$$

marginal distribution.

$$\int p(A, B) dB = P(A) \leftarrow$$

⑤ posterior predictive $\approx$ parametric model (true).

⑤ parameter $\approx$ $\approx$ $\approx$ $\approx$ (false).

예제 2

주어진 관측.

$x_1, x_2, \dots, x_{100}$

Gaussian model.
 $\sim N(\theta, 1^2)$

평균 θ 를 추측하고 싶음.

① likelihood $\rightarrow$ MLE.

$$\hat{\theta}_{MLE} = \frac{1}{100} \sum_{n=1}^{100} x_n \leftarrow \text{평균}$$

$$\mathcal{L}(\theta) = p(x_1, x_2, \dots, x_{100}; \theta)$$

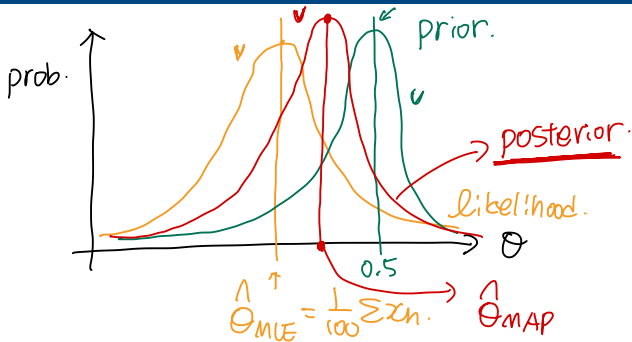
$$= \prod p(x_n; \theta).$$

$$\log \mathcal{L}(\theta) = \sum \log p(x_n; \theta).$$

$$\frac{d}{d\theta} \log \mathcal{L}(\theta) = 0$$

$$\hat{\theta}_{MLE} = \operatorname{argmax} \mathcal{L}(\theta)$$

$$= \operatorname{argmax} \log \mathcal{L}(\theta)$$



② posterior $\rightarrow$ MAP.

Θ prior distribution.

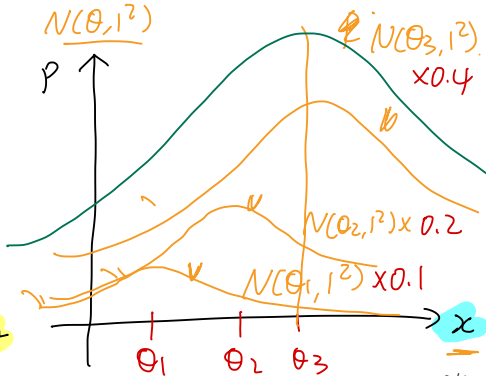
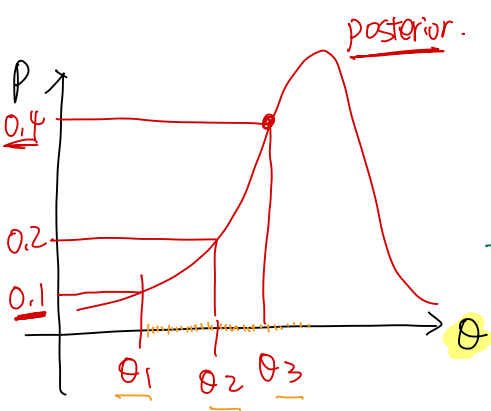
$$p(\theta) = \mathcal{N}(0.5, 0.05^2)$$

$$p(\theta | x_1, \dots, x_N) \propto \underbrace{p(x_1, x_2, \dots, x_N)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

③ posterior predictive.

$p(\theta | x_1, x_2, \dots, x_N) \leftarrow$ 앞 데이터로 얻은 θ 의 분포.

$$p(x | x_1, x_2, \dots, x_N) = \int p(x | \theta) p(\theta | x_1, x_2, \dots, x_N) d\theta.$$



How to treat $p(x' \mid x_1, x_2, \dots, x_N)$?

- Maximum likelihood estimate

$$\hat{\theta}_{\text{MLE}} = \arg \max p(x_1, x_2, \dots, x_N \mid \theta) \quad (1)$$

$$p(x' \mid x_1, x_2, \dots, x_N) \sim p(x' \mid \hat{\theta}_{\text{MLE}}) \quad (2)$$

- Maximum a posteriori estimate

$$\hat{\theta}_{\text{MAP}} = \arg \max p(\theta \mid x_1, x_2, \dots, x_N) \quad (3)$$

$$\text{where } p(\theta \mid x_1, x_2, \dots, x_N) \propto p(x_1, x_2, \dots, x_N \mid \theta)p(\theta) \quad (4)$$

$$p(x' \mid x_1, x_2, \dots, x_N) \sim p(x' \mid \hat{\theta}_{\text{MAP}}) \quad (5)$$

How to treat $p(x' \mid x_1, x_2, \dots, x_N)$?

- Posterior predictive distribution

$$p(x' \mid x_1, x_2, \dots, x_N) = \int_{\theta} p(x' \mid \theta) p(\theta \mid x_1, x_2, \dots, x_N) d\theta \quad (6)$$