likelihood $\Leftarrow$ ① $p(x;\theta)$

관측 확률

~~θ의 prob~~

$x_1, x_2, \cdots, x_N$

$$\hat{\theta} = \operatorname{argmax} \text{②} \mathcal{L}(\theta) = \operatorname{argmax} \text{③} p(x_1, x_2, x_3 \cdots x_N; \theta)$$

Inclass 21: Maximum A Posteriori Estimate

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

	H	T
θ_1	0.5	0.5
θ_2	0.9	0.1

관측

TT

$\mathcal{L}(\theta_1) = 0.25 \Leftarrow$ 관측될 확률

~~θ_1일 확률이 0.25다.~~

$\mathcal{L}(\theta_2) = 0.01$

$$\mathcal{L}(\theta) = Pr(x_1, x_2 \cdots, x_N; \theta)$$

Pr		H	T
0.1	→ θ_1	0.5	0.5 =
0.9	→ θ_2	0.9	0.1

θ_2.

관측

T T T T T . . . .

↓

θ_2 · · ··· θ_1?

		H	T
$0.1 \rightarrow \theta_1$		0.5	0.5
$0.9 \rightarrow \theta_2$		0.9	0.1

관측

T T

$$\left\{ p(\theta_1) = 0.1 \right\} \times \left\{ \mathcal{L}(\theta_1) = P(TT;\theta_1) = 0.25 \right\} = \left\{ p(\theta_1 \mid T) \right\} \propto 0.025$$

$$\left\{ p(\theta_2) = 0.9 \right\} \times \left\{ \mathcal{L}(\theta_2) = P(TT;\theta_2) = 0.01 \right\} = \left\{ p(\theta_2 \mid T) \right\} \propto 0.009$$

prior

posterior

$$\hat{\theta}_{MAP} = \operatorname{argmax} \ P(\theta \mid x_1, x_2) = \theta_1$$

$$\hat{\theta}_{MLE} = \operatorname{argmax} \ P(x_1, x_2 \mid \theta) = \theta_1$$

Bayes' theorem

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

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

posterior

$$p(\theta|D) \propto p(D|\theta)\,p(\theta)$$

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

$D = x_1, x_2, \cdots, x_N$

관측이 주어진 이후의 θ의 확률분포.

관측이 주어지기 이전에 θ에 대해 가지고 있는 확률분포.

Maximum a posteriori estimate

- Maximum likelihood estimate

$$\hat{\theta}_{\text{MLE}} = \arg\max \mathcal{L}(\theta) \tag{1}$$
$$= \arg\max p(x_1, x_2, ..., x_N \mid \theta) \tag{2}$$

- Maximum a posteriori estimate (MAP estimate)

$$\hat{\theta}_{\text{MAP}} = \arg\max p(\theta \mid x_1, x_2, ..., x_N) \tag{3}$$
$$= \arg\max p(x_1, x_2, ..., x_N \mid \theta) p(\theta) \tag{4}$$

likelihood $\Leftarrow$ $p(x;\theta)$

MLE

$\theta \sim Pr(\theta)$

posterior $\Leftarrow$ $p(x;\theta)$, $p(\theta)$

MAP.

θ가 random variable.

$p(\theta)$, $p(\theta \mid x_1, x_2 \cdots, x_N)$

ⓥ $\hat{\theta}_{MAP} = \hat{\theta}_{MLE},$

$\underline{\theta}$ ← prior prob가 uniform 일때.

prior.

ⓥ MLE 은, uniform θ의 MAP와 동일하다.

= θ에 대한

어떠한 선입견 X

$$\begin{array}{lccc} & & H & T \\ 0.5 \rightarrow & \theta_1 & 0.5 & 0.5 \\ 0.5 \rightarrow & \theta_2 & 0.9 & 0.1 \end{array}$$

TT

$0.5 \; \underline{P(\theta_1)} \times \underline{\mathcal{L}(\theta_1) = 0.25} \propto \underline{P(\theta_1 | TT)}$

$0.5 \; \underline{P(\theta_2)} \times \underline{\mathcal{L}(\theta_2) = 0.01} \propto \underline{P(\theta_2 | TT)}$

$\hat{\theta} = \operatorname{argmax} \mathcal{L}(\theta) \stackrel{?}{=} \hat{\theta} = \operatorname{argmax} P(\theta | D)$

	H	T
$0.01 \to \theta_1$	0.5	0.5
$0.99 \to \theta_2$	0.9	0.1

관측

T T T ......

$\hat{\theta}_{MAP} = \theta_2 \to \theta_1$

	T	TT	TTT	TTTT..	
관측개수	1	2	3	4	. . .
$P(\theta_1) \times \mathcal{L}(\theta_1)$	$\frac{5}{10}$	$\frac{25}{100}$	$\frac{125}{1000}$	$\frac{625}{10000}$	$\times 0.01$ (100배)
$P(\theta_2) \times \mathcal{L}(\theta_2)$	$\frac{1}{10}$	$\frac{1}{100}$	$\frac{1}{1000}$	$\frac{1}{10000}$	$\times 0.99$
MAP	θ_2	θ_2	θ_1	θ_1	