

model + cost $\rightarrow$ opt.
linear SSE normal
G.D. $\rightarrow$ 방법론

Inclass 15: MLE and MAP : Statistical inference.

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.ac.kr)

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

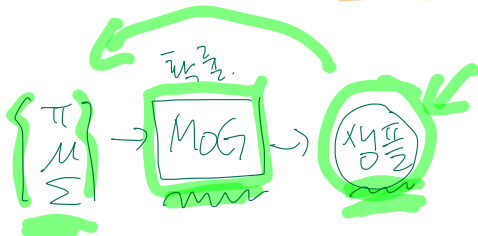
이론

최적화

2차식

선형

$\frac{1}{2}x^2 + \frac{1}{2}x$



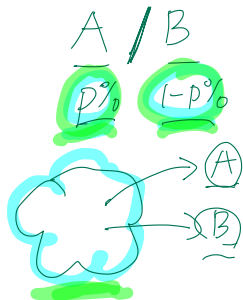
prob.

$$\frac{1}{2} \times \frac{2}{2} =$$

$$\frac{1}{2} \times \frac{2}{2} =$$

stat.

$$\frac{1}{2} \times \frac{2}{2} =$$



기초

/

분류

/

심화

SVM

→

convex

L.R.

G.D.

back

perceptron → neural network.

decision tree.

k-means

MLE

MAP

k-nn

MoG.

MLE = Maximum Likelihood Estimate

① ⁶ Likelihood? ⁹

② ⁶ Max. ⁹ Likelihood Estimate?

공정성. ←
Gaussian.

θ_1

	H	T
θ_1 pr	0.9	0.1
θ_2 pr	0.1	0.9

관측
사건
샘플
0.2

①	②	③	④	⑤	⑥	...
H	H	H	T	T	T	T

likelihood $\rightarrow$ parameter

$$\mathcal{L}(\theta_1) = \text{Pr}(HHH; \theta_1) = \frac{0.9 \times 0.9 \times 0.9}{0.001}$$

$$\mathcal{L}(\theta_2) = \text{Pr}(HHH; \theta_2) = \frac{0.1 \times 0.1 \times 0.1}{0.001}$$

$$\underline{\mathcal{L}(\theta)} = \underline{Pr}(\underline{x_1, x_2, \dots, x_N} \overset{\phi}{;} \underline{\theta})$$

$\Rightarrow \underline{\theta}$ 는 R.V 이든 아니든 상관 X

$$\Rightarrow \underline{Pr(X ; \theta)}$$

X^2 을 관측하는 $\overline{X^2}$ 이

θ 의 영향을 받음.

$\overline{X^2}$ 을 보면 $\rightarrow$ θ 가짐.

Maximum likelihood estimate

Maximum likelihood estimation, or MLE, is on flavor of parameter estimation in machine learning. In order to perform parameter estimation, we need:

- some data $\mathbf{x}$
- some hypothesized generating function of the data $f(\mathbf{x}, \theta)$
- a set of parameters from that function θ
- some evaluation of the goodness of our parameters

Maximum likelihood estimate

In MLE, the objective function (evaluation) we chose is the likelihood of the data given our model. To find the best θ then, we need to find the θ which maximizes our evaluation function (the likelihood).

Therefore, in its general form the MLE is:

$$\begin{aligned}\hat{\theta}_{\text{MLE}} &= \arg \max_{\theta} \mathcal{L}(\theta) = \Pr(x_1, x_2, \dots, x_N; \theta)_{(1)} \\ &= \arg \max_{\theta} \log \mathcal{L}(\theta) = \prod \Pr(x_n; \theta)_{(2)}\end{aligned}$$

$$\theta \rightarrow \mathcal{L}(\theta) \rightarrow ?$$

$\uparrow$
 $x_1 \dots x_N$

	H	T
θ	θ	$1-\theta$

HHT



$$L(\theta) = \Pr(HHT; \theta) = \theta^2(1-\theta)$$

$$\log L(\theta) = 2 \log \theta + \log(1-\theta)$$

$$\frac{d}{d\theta} (\cdot) = \frac{2}{\theta} - \frac{1}{1-\theta} = 0$$

$$\frac{2}{\theta} = \frac{1}{1-\theta}$$

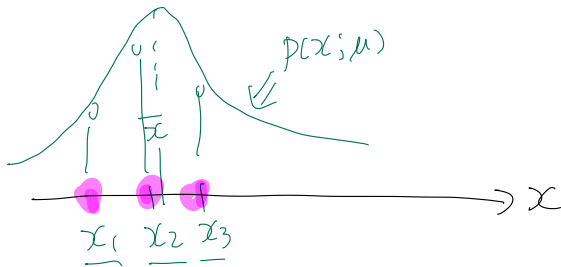
$$2 - 2\theta = \theta$$

MLE solution
↓

$$\hat{\theta} = \frac{2}{3}$$

Gaussian distribution.

$$\theta = \mu, \quad \underline{\underline{\sigma^2 = 1}}$$



$$\underline{x_1}, \underline{x_2}, \underline{x_3}.$$

$$\underline{\hat{\mu} = \frac{1}{3} (x_1 + x_2 + x_3)}$$

$$x_1, x_2, x_3, \quad \theta = \mu, \quad \sigma^2 = 1.$$

$$p(\underline{x}; \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ - \frac{(\underline{x} - \theta)^2}{2} \right\} \ll$$

$$\mathcal{L}(\theta) = p(x_1; \theta) p(x_2; \theta) p(x_3; \theta)$$

$$\log \mathcal{L}(\theta) = \sum_{n=1}^3 \log p(x_n; \theta)$$

$$= \sum_{n=1}^3 \left(-\frac{1}{2} \log 2\pi - \frac{(x_n - \theta)^2}{2} \right)$$

$$\frac{d}{d\theta} \log \mathcal{L}(\theta) = \sum_{n=1}^3 (x_n - \theta) = 0.$$

$$\Rightarrow \frac{1}{3} \sum_{n=1}^3 x_n = \hat{\theta}_{MLE}$$


$$\theta \rightarrow \underline{L(\theta)}$$

$$= \text{Pr}(x_1 \cdots x_N; \theta)$$

$$\uparrow$$

$$x_1 \cdots x_N$$

$$\underline{\text{Pr}(x; \theta)}$$

$$\hat{\theta}_{\text{MLE}} = \arg \max L(\theta)$$

$$= \arg \max \log L(\theta)$$

Maximum a posteriori estimate

- Maximum likelihood estimate

$$\leftarrow \underline{p(x; \theta)}$$

$$\hat{\theta}_{\text{MLE}} = \arg \max \mathcal{L}(\theta) \quad (3)$$

$$= \arg \max p(x_1, x_2, \dots, x_N | \theta) \quad (4)$$

- Maximum a posteriori estimate

$$\underline{p(x; \theta)} \oplus \underline{p(\theta)}$$

$$\hat{\theta}_{\text{MAP}} = \arg \max p(\theta | x_1, x_2, \dots, x_N) \quad (5)$$

$$= \arg \max \underline{p(x_1, x_2, \dots, x_N | \theta)} \underline{p(\theta)} \quad (6)$$

Bayesian approach.
(non-parametric)

$\Rightarrow$ random variable

$$\underline{p(\theta)}, \quad \underline{p(\theta | D)}$$

$\uparrow \quad \quad \uparrow$

MLE: $p(x; \theta)$

$$\mathcal{L}(\theta) = \text{pr}(\underbrace{x_1, x_2, \dots, x_N}_{\downarrow}; \theta)$$

MAP: maximum a posteriori estimate

posterior ↙

$$P(\theta | x) = \frac{p(x|\theta)p(\theta)}{\sum_{\theta'} p(x|\theta')p(\theta')}$$

$$\frac{p(x|\theta)p(\theta)}{P(x)}$$

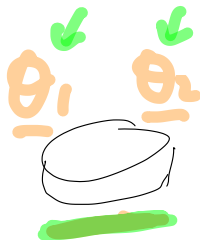
Bayes' theorem.

$$= \frac{p(x|\theta)p(\theta)}{\sum_{\theta'} p(x|\theta')p(\theta')}$$

likelihood prior

$$\alpha \cdot p(x|\theta) p(\theta)$$

	H	T	
θ_1	0.9	0.1	←
θ_2	0.1	0.9	←

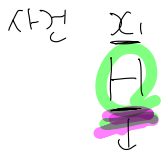


	θ_1	θ_2	
θ	0.1	0.9	←

$p(\theta)$

prior
↓
 θ

	H	T
θ_1	0.9	0.1
θ_2	0.1	0.9



$$\mathcal{L}(\theta_1) = 0.9$$

$$\mathcal{L}(\theta_2) = 0.1$$

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



$$\mathcal{L}(\theta_1) = 0.81$$

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

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

	θ_1	θ_2
θ	$\frac{1}{101}$	$\frac{100}{101}$

$$p(\theta_1) = \frac{1}{101}$$

$$p(\theta_1 | HH)$$

$$\propto \frac{p(HH|\theta_1)}{p(HH)} p(\theta_1)$$

$$\frac{0.1}{100} \times \frac{1}{101}$$

$$p(\theta_2 | HH)$$

$$\propto \frac{p(HH|\theta_2)}{p(HH)} p(\theta_2)$$

$$\frac{1}{100} \times \frac{100}{101}$$

$$\underline{p(\theta_1 | HH)} \propto \frac{81}{100} \times \frac{1}{101} = \frac{81 \leq}{10100}$$

$$\underline{p(\theta_2 | HH)} \propto \frac{1}{100} \times \frac{100}{101} = \frac{100 \leftarrow}{10100}$$

$$p(\underline{\theta_1} | \underline{HH}) = \frac{81}{181} \nwarrow$$

$$p(\underline{\theta_1}) = \frac{1}{101} \nearrow$$

$$p(\underline{\theta_2} | \underline{HH}) = \frac{100}{181} \nwarrow$$

$$p(\underline{\theta_2}) = \frac{100}{101} \searrow$$

	H	T
θ_1	0.9	0.1
θ_2	0.1	0.9

	θ_1	θ_2
θ	$\frac{1}{101}$	$\frac{100}{101}$

x_1
H

x_2
H

x_3
H

$\mathcal{L}(\theta_1)$	0.9	0.01	0.729
$\mathcal{L}(\theta_2)$	0.1	0.01	0.001
$P(\theta_1 D)$	$\frac{9}{10} \times \frac{1}{101}$	$\frac{0.1}{100} \times \frac{1}{101}$	$\frac{729}{1000} \times \frac{1}{101}$
$P(\theta_2 D)$	$\frac{1}{10} \times \frac{100}{101}$	$\frac{1}{100} \times \frac{100}{101}$	$\frac{1}{1000} \times \frac{100}{101}$
$\hat{\theta}_{MLE}$	θ_1	θ_1	θ_1
$\hat{\theta}_{MAP}$	θ_2	θ_2	θ_1