기말고사 12/07 (수)

샘플들

• 통계적추론: $x_1, x_2, x_3, x_4, \cdots x_N \longrightarrow \theta$ 측정.

최대 우도 측정

Inclass 20: Maximum Likelihood Estimate (MLE)

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

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

개론, 기초

linear regression,

logistic regression

perceptron

k-nearest neighbor.

머신러닝

gradient descent

decision tree

random forest

k-means

moG.

PCA

심화

support vector machine

information

convex opt.

MLE

MAP

미분

최적화.

머신러닝

선형대수

확률통계

6주간?

① Likelihood 란 무엇인가?

② maximum Likelihood estimate

예

① 동전 biased coin

2개

3개

real value

② Gaussian 분포, mean 추정.

$\theta = \theta_1$

	H	T
pr	0.5	0.5

$Pr(X=H;\theta_1)=0.5$

샘플수집.

T T T H

$\mathcal{L}$: 어떤 사건이 얼마나 일어날법한가.

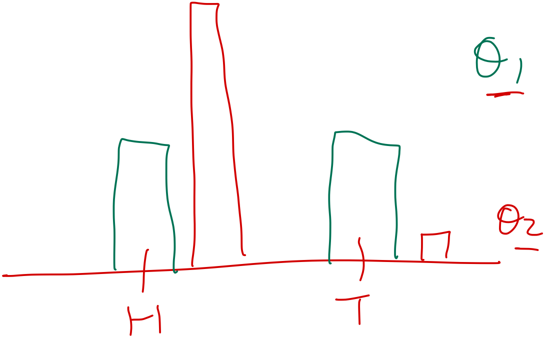
이 관측이 실제로 관측이 될 확률이 얼마냐.

$\theta = \theta_2$

	H	T
pr	0.9	0.1

$Pr(X=H;\theta_2)=0.9$

$\mathcal{L}(\theta)$



θ_1
θ_2
H
T

$\mathcal{L}(\theta=\theta_1) = \Pr(TTTH; \theta=\theta_1)$

1번 동전일때,
TTTH가
관측될 확률.

$= \dfrac{625}{10000}$

$\mathcal{L}(\theta=\theta_2) = \Pr(TTTH; \theta=\theta_2)$

2번 동전일때,
TTTH가
" .

$= \dfrac{9}{10000}$

θ의 함수. → $\mathcal{L}(\theta)$

샘플들, 관측들. →

$= \Pr(x_1, x_2 \cdots x_N ; \theta)$

$\Pr(x ; \theta)$

관측될 확률 분포가
θ의 영향을 받고,
θ가 주어져야
계산이 가능.

	H	T
θ_1	0.5	0.5
θ_2	0.9	0.1
θ_3	0.1	0.9

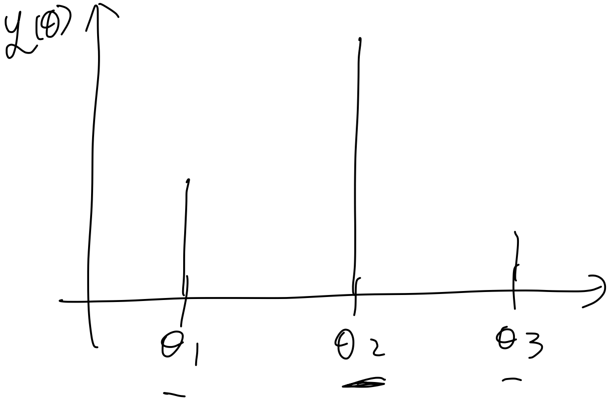
$x_1\ x_2\ x_3\ x_4$

HHHT

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; x_1, x_2, x_3, x_4)$$

$$\begin{cases} \mathcal{L}(\theta_1) = 125/10000 \\ \mathcal{L}(\theta_2) = 729/10000 \\ \mathcal{L}(\theta_3) = 9/10000 \end{cases}$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\arg\max}\ \mathcal{L}(\theta) = \theta_2$$



$\mathcal{L}(\theta)$
θ_1
θ_2
θ_3

H	T
θ	$1-\theta$

$\Pr(X=H;\theta)=\theta$

$\Pr(X=T;\theta)=1-\theta$

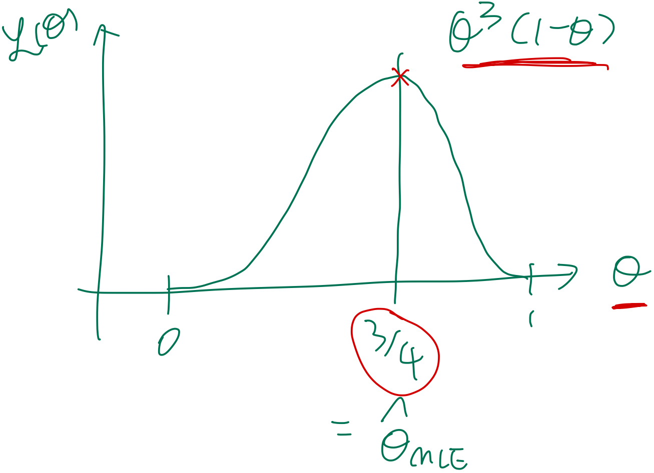
$\theta\in[0,1]$

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ H & H & H & T \end{pmatrix}$$

$$\mathcal{L}(\theta)=\Pr(x_1=H,\ x_2=H,\ x_3=H,\ x_4=T;\theta)$$

$$=\Pr(x_1=H;\theta)\Pr(x_2=H;\theta)\Pr(x_3=H;\theta)\Pr(x_4=T;\theta)$$

$$=\theta^3(1-\theta)$$


$\mathcal{L}(\theta)$
$\theta^3(1-\theta)$
θ
0
$3/4$
1
$= \hat{\theta}_{MLE}$

$$\hat{\theta}_{MLE} = \underset{\sim}{\arg\max}\ \underline{\mathcal{L}(\theta)}$$

$$= \underline{\arg\max}\ \underline{\log}\ \underline{\mathcal{L}(\theta)}$$

$$\underline{\log \mathcal{L}(\theta)} = \log \theta^3 (1-\theta)$$

$$= \underline{3\log\theta + \log(1-\theta)}$$

$$\frac{d}{d\theta} \log \mathcal{L}(\theta) = \frac{3}{\theta} - \frac{1}{1-\theta} = 0$$

$$\theta = \frac{3}{4}$$

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 $\mathbf{x}$ $(x_1, x_2, \cdots, x_N)$
- 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 = Likelihood

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:

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} \mathcal{L}(\theta) \quad (1)$$

$$= \arg\max_{\theta} \log \mathcal{L}(\theta) \quad (2)$$

MoG.

$\pi_k, \mu_k, \Sigma_k \leftarrow \gamma_{nk}$

$\Rightarrow \pi_k = \frac{N_k}{N} \quad \mu_k = \frac{1}{N_k} \sum \gamma_{nk} x_n \quad \ldots$

μ를 추정.

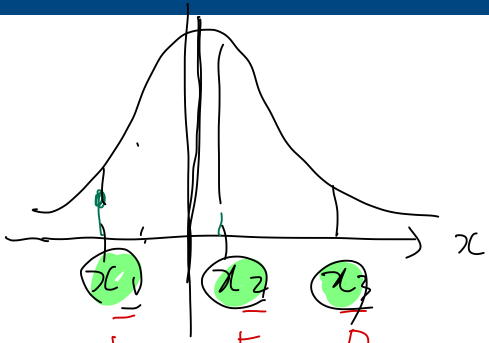
$$x \sim N(\mu, 1)$$

$$x_1, x_2, \cdots x_5$$

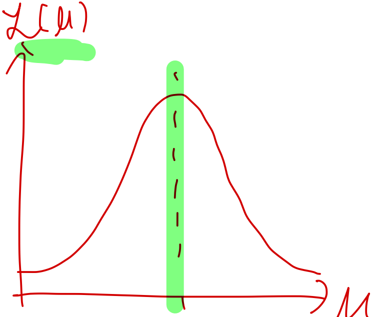
$$\hat{\mu} = \frac{1}{5}(x_1 + x_2 + \cdots x_5)$$

'likelihood'

x_1, x_2, x_3



$\bar{x} = \hat{\mu}_{MLE} = \frac{13}{3}$



$\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3)$

$\mathcal{L}(\mu_1) =$

$\mathcal{L}(\mu_2) =$

$$\mathcal{L}(\mu) = \Pr(x_1, x_2, x_3 ; \mu)$$

$$= \Pr(x_1 ; \mu) \Pr(x_2 ; \mu) \Pr(x_3 ; \mu)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_1-\mu)^2}{2}\right)$$

$$\times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_2-\mu)^2}{2}\right)$$

$$\times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_3-\mu)^2}{2}\right)$$

$$\log \mathcal{L}(\theta) = \cancel{3\log\sqrt{2\pi}} - \frac{(x_1-\mu)^2}{2} - \frac{(x_2-\mu)^2}{2} - \frac{(x_3-\mu)^2}{2}$$

$$\frac{d}{d\mu}\log\mathcal{L}(\mu) = (x_1-\mu) + (x_2-\mu) + (x_3-\mu) = 0$$

$$= 0$$

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