

# Model Selection : MAP, MLE

## Introduction

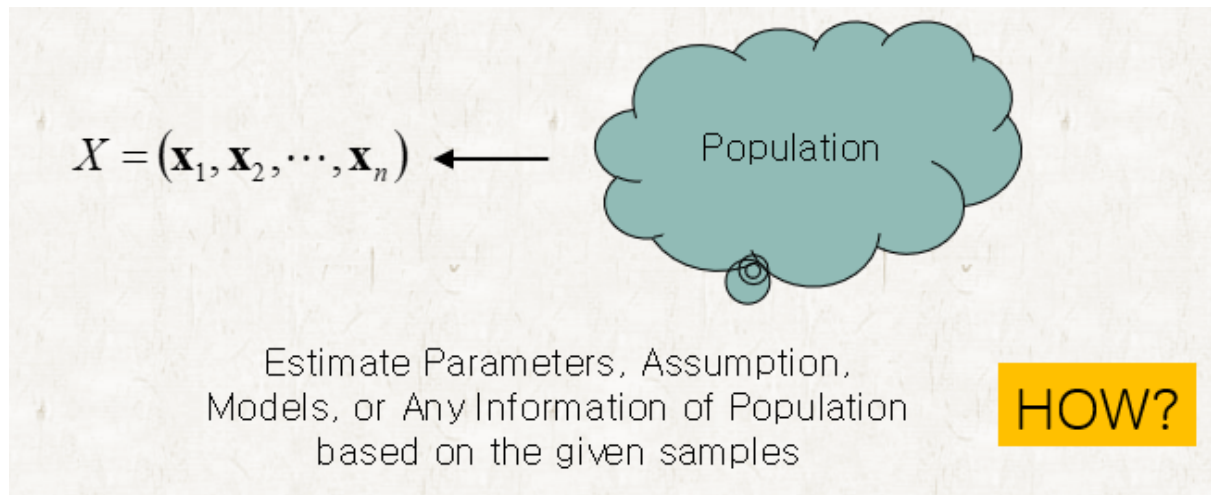
### How to find a model for given data

- Based on **Error Function**
  - Define an error function
  - Find a model which **minizes the error function**
    - solving linear equation
    - Gradient descent method
- Based on Probability
  - Evaluate the probability of given data generated
  - Find a model which **maximize the probability**
    - Solving linear equation
    - Gradient descent method
    - EM(Expectation-Maximization)
    - MCMC(Markov Chain Monte Carlo)

## Model Selection

### You have sample data $x=(x_1, x_2, \dots, x_n)$

- Do not know about population.
- want to estimate some hidden parameters of population >> model
- sample data가 있는데 이걸 어떤 모델에서 나올까??



### Let's assume that we have

- a set of sample data

$$D = \{d_1, d_2, \dots, d_n\}$$

- a set of candidate models, parameters or assumptions

$$M = \{m_1, m_2, \dots, m_k\}$$

D를 가장 잘 설명하는 모델m을 골라야한다.>>p(m|D)

### Usually

- $P(D|mi)$  is easy to evaluate, but  $P(mi|D)$  is hard

### Questions

- Choose the most probable model given the training data

### Choose a model by Maximum A Poerior(MAP)

(Choose the most probable model given the training data)

- Evaluate the conditional probability of each  $m$  given  $D$

$$P(m_1 | D), P(m_2 | D), \dots, P(m_k | D)$$

- Choose the model which has the maximum probability
- In short

$$\text{model} = \arg \max_{m \in M} P(m | D)$$

$$\begin{aligned} \text{model} &= \arg \max_{m \in M} P(m | D) \\ &= \arg \max_{m \in M} \frac{P(D | m)P(m)}{P(D)} \\ &= \arg \max_{m \in M} \left\{ \frac{P(D | m_1)P(m_1)}{P(D)}, \dots, \frac{P(D | m_k)P(m_k)}{P(D)} \right\} \\ &= \arg \max_{m \in M} P(D | m)P(m) \end{aligned}$$

• Also, remember

$$\arg \max_{m \in M} P(m | D) = \arg \max_{m \in M} \log P(m | D)$$

그런데 왜  $p(m)$ 은 구하기 어려운가?

예를들어 모델이 10개가 있다치면 1/10이아닌가?

>>단정지을수 없다.

## Maximum Likelihood Estimator(MLE)

- Maximum Likelihood Estimator (MLE)

$$\text{model} = \arg \max_{m \in \mathcal{M}} P(D | m) P(m)$$

Assume that  $P(m)$  is the same for all  $m$

$$= \arg \max_{m \in \mathcal{M}} P(D | m)$$

If  $m$  is given, it is Probability of D  
If D is given, it is Likelihood of m

- Choose a model which maximizes the likelihood

$$\text{model} = \arg \max_{m \in \mathcal{M}} P(D | m)$$

MAP에서  $P(m)$ 을 다 동일하다 가정 즉 모델이  $k$ 개 있으면  $p(m)=1/k$

그러면  $p(m)$ 은 상수가 되니까

$\arg \max P(D|m)$ 만 구하면 된다 << 이것이 MLE

보통 MLE를 구하는 것이 MAP보다 쉽다.

👁 **likelihood**:  $p(z|x)$ , 어떤 모델에서 해당 데이터(관측값)이 나올 확률

👁 **사전확률(prior probability)**:  $p(x)$ , 관측자가 관측을 하기 전에 시스템 또는 모델에 대해 가지고 있는 선형적 확률. 예를 들어, 남녀의 구성비를 나타내는  $p(\text{남자})$ ,  $p(\text{여자})$  등이 사전 확률에 해당한다.

👁 **사후확률(posterior probability)**:  $p(x|z)$ , 사건이 발생한 후(관측이 진행된 후) 그 사건이 특정 모델에서 발생했을 확률

- Evaluate the conditional probability of D given each of m

$$P(D | m_1), P(D | m_2), \dots, P(D | m_k)$$

- Choose the model which has the maximum probability

- In short

$$\text{model} = \arg \max_{m \in M} P(D | m)$$

$$P(h | c_1) = 1.0, P(h | c_2) = 0.5, P(h | c_3) = 0.0,$$

## Example - Coin Tossing

- Example: Coin Tossing
  - We have five coins in a pockets, and choose one randomly and toss it
    - One of  $c_1$ : two heads
    - Three of  $c_2$ : one head and one tail
    - One of  $c_3$ : two tails
  - Now, we see a head. Which coin do you chose?
    - What is the Sample?  $D = \{h\}$
    - What are the candidates of what to estimate?  $M = \{c_1, c_2, c_3\}$
  - MAP estimator of coin

$$\begin{aligned} \text{model} &= \arg \max_{c \in \{c_1, c_2, c_3\}} P(c | D) \\ &= \arg \max_{c \in \{c_1, c_2, c_3\}} P(c | h) \\ &= \arg \max_{c \in \{c_1, c_2, c_3\}} P(h | c) P(c) \end{aligned}$$

$$P(c_1) = 0.2, P(c_2) = 0.6, P(c_3) = 0.2$$

$$P(h|c_1)=1.0, P(h|c_2)=0.5, P(h|c_3)=0.0,$$

$$\left. \begin{array}{l} P(h|c_1)P(c_1)=0.2 \\ P(h|c_2)P(c_2)=0.3 \\ P(h|c_3)P(c_3)=0.0 \end{array} \right\} \xrightarrow{\max} P(h|c_2)P(c_2)=0.3$$

MAP estimator is  $c_3$