PRML Chapter.1

# Introduction

2017010698
수학과 오서영

# 목차

# 2. Probability Theory

**패턴인식 -> 불확실성(uncertainty) 의 문제**

: finite size of data sets, Noise on measurements

**- Probability Theory**

: consistent framework
for the quantification and manipulation of uncertainty

**- Decision Theory**

: incomplete or ambiguous information -> optimal predictions

# 2. Probability Theory

## Probability Theory – Frequentist

: 실제 데이터가 존재해야 불확실성을 정량화 가능
-> 빈도를 통해 모델링

## Probability Theory - Bayesian probabilities

: 사건이 발생하지 않은 경우에도 확률을 부여할 수 있다
-> 좀 더 불확실한 경우에도 모델링이 가능

Curve Fitting 문제에서는 모수 w가 알려지지 않은
고정된 값으로 여겨졌다. (unknown but fixed)
하지만 베이지안 방식은 모든 값이 확률 값이므로
이를 하나의 확률 변수로 고려

# 2. Probability Theory

## Probability Theory - Bayesian probabilities

$$p(\mathbf{w}|D) = \frac{p(D|\mathbf{w})p(\mathbf{w})}{p(D)}$$

- 관찰되는 데이터가 존재하기 이전에 이미 p(w)를 통해
w 의 불확실한 정도를 수식에 반영
-> prior probability distribution **p(w)**

- how probable the observed data set
is for different settings of the w
-> likelihood **p(D|w)**

$$posterior \propto likelihood \times prior$$

normalization $p(D) = \int p(D|\mathbf{w})p(\mathbf{w})d\mathbf{w}$

# 2. Probability Theory

## 가우시안 분포 (Gaussian distribution) = 정규분포

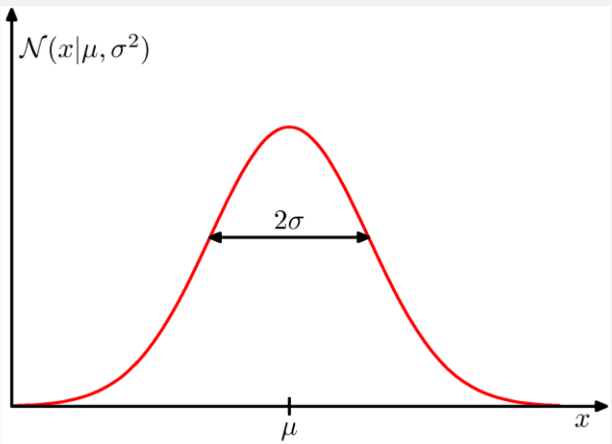
$$N(x \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

**Parameter** :

μ : mean

σ : standard deviation

**정확도(precision)** := 1/σ^2




$$N(x \mid \mu, \sigma^2) > 0 \qquad (1.47)$$

$$\int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2)\, dx = 1 \qquad (1.48)$$

$$E[x] = \int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) \cdot x \, dx = \mu \qquad (1.49)$$

$$E[x^2] = \int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) \cdot x^2 \, dx = \mu^2 + \sigma^2 \qquad (1.50)$$

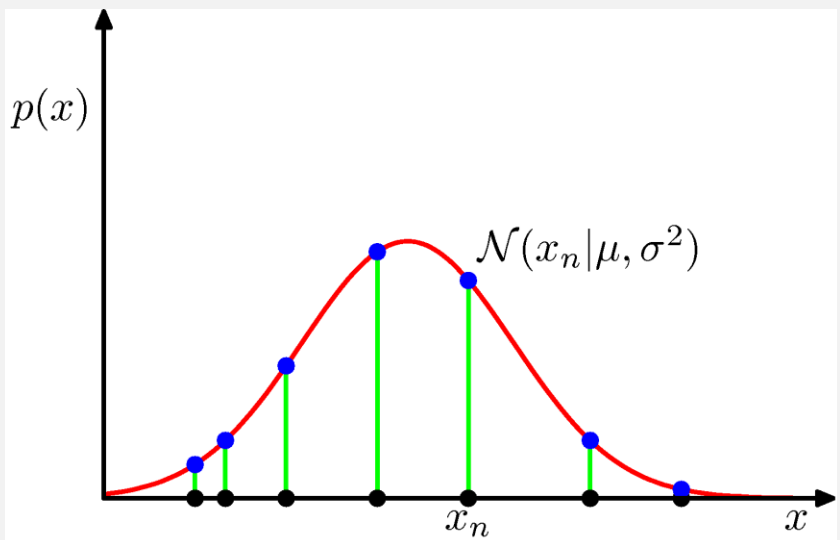
$$var[x] = E[x^2] - E[x]^2 = \sigma^2 \qquad (1.51)$$

# 2. Probability Theory

(EX) 관찰 데이터 집합 x=(x1,...,xN).T
-> 이 데이터 집합 하나가 관찰될 수 있는 확률?

각각의 데이터가 발현되는 가능성 : i.i.d

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} N(x_n \mid \mu, \sigma^2)$$



# 2. Probability Theory

(EX) 관찰 데이터가 하나의 가우시안 분포를 따름.
얻는것 : 관찰 데이터 집합
문제 : 가우시안 분포를 결정
-> 주어진 샘플이 어떤 가우시안 분포에서 나왔는지 결정
-> p(x | μ, σ^2) 를 이용하여 이러한 관찰 결과를
만들어낼 만하다고 생각할 수 있는 가장 타당한 μ 와 σ 찾기

→ **Parameter estimation**

$$\ln p(\mathbf{x} \mid \mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

→ **maximize**

각각 샘플에 대한
평균, 분산

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad (1.55)$$

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 \qquad (1.56)$$

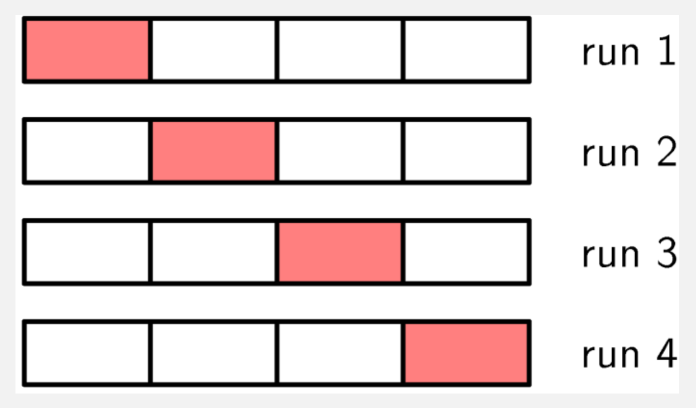
→ *MLE* (*Maximum likelihood estimation*) -> 쉽게 overfit

# 3. Model Selection

데이터가 충분하다면 다양한 모델을 학습하여
가장 적당한 파라미터 값을 추정할 수 있다

**cross validation**

: Random 으로 고른 (S−1)/S 를 Training Set 으로,
1/S 를 Test set 으로 학습 진행
**->** 학습 데이터에만 종속적이고
Overfitting이 발생하지 않는 측정값을 찾는 것.



# 3. Model Selection

**infomation criteria**

: MLE 로 인해 발생되는 overfit을 막기 위해 페널티 조건을 추가

(EX) **AIC**(Akaike information criterion)

$$\ln p(D|\mathbf{w}_{ML}) - M$$

**M** : 모델에서 사용한 parameter 의 개수 (penalty)

현재 주어진 데이터를 가장 적합하게 만드는 parameter 를 추정하되 모델이 복잡해서는 안된다

즉, 동일한 성능을 내는 모델 두개가 제공된다면 여기서 덜 복잡한 모델을 선택할 수 있어야 한다.