

START

머신러닝과 딥러닝

Machine Learning & deep Learning

Chapter 5. 선형분류모형 II

Machine Learning & Deep Learning

손영두

e-mail: youngdoo@dongguk.edu

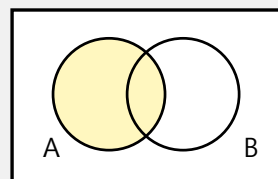


조건부 확률과 베이즈 정리

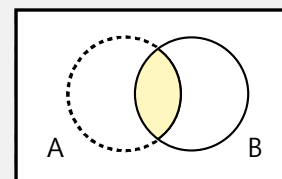
조건부 확률 (Conditional Probability)

- ✓ 특정 사건이 벌어졌다는 가정 하에서 또 다른 사건이 벌어질 확률
- ✓ B가 주어진 상황에서 A의 조건부 확률

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



(a)



(b)

- ✓ 조건부 확률 $P(A|B)$ 는 확률의 정의를 모두 만족

$$P(A|B) \geq 0$$

$$P(S|B) = 1$$

$$A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$$

- ✓ 주어진 상태 B를 조건 혹은 사전 정보(prior information)라고 한다



조건부 확률과 베이즈 정리

■ 조건부 확률 (Conditional Probability)

☑ 결합 확률 (joint probability)

- 임의로 뽑힌 여자 한 명이 정시 입학한 여학생일 확률은?
- 여자(F)와 정시(R) 사건이 동시에 발생: $P(F \cap R) = 15/60$

☑ 주변 확률 (marginal probability)

- 뽑힌 학생이 수시 입학한 학생일 확률은?
- $P(E) = 25/60$

	남자(M)	여자(F)	합계
수시(E)	12	13	25
정시(R)	20	15	35
합계	32	28	60

☑ 조건부 확률 (conditional probability)

- 뽑힌 학생이 남학생임을 알 때, 이 남학생이 정시 입학한 학생일 확률은?
- 남학생 : 32명, 남학생 중 정시 입학 : 20명
: $20/32$



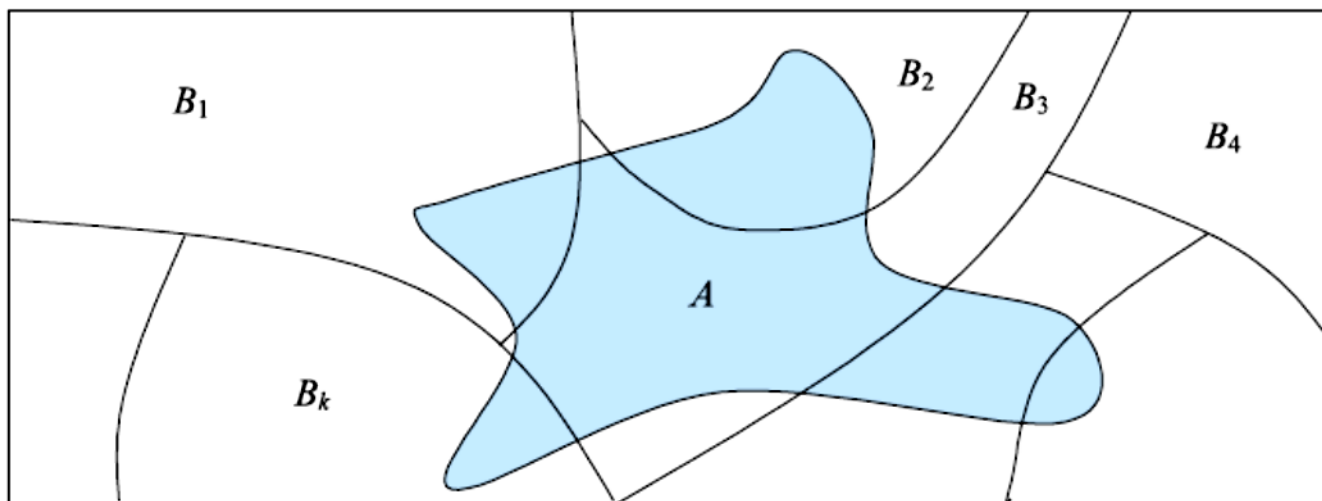
조건부 확률과 베이즈 정리

■ 베이즈 정리

전확률 정리(theorem of total probability)

표본공간 S 를 서로 배반인 사상 B_1, B_2, \dots, B_k 로 분할하였을 때, S 의 임의의 사

상 A 에 대하여 $P(A) = \sum_{i=1}^k P(B_i \cap A) = \sum_{i=1}^k P(B_i)P(A|B_i)$ 가 성립한다.





조건부 확률과 베이즈 정리

■ 베이즈 정리

베이즈 정리(Bayes Theorem)

사상 B_1, B_2, \dots, B_k 가 표본공간 S 의 분할이고, $P(B_i) \neq 0, i = 1, 2, \dots, k$ 이고, 임의의 사상 $A(P(A) \neq 0)$ 에 대하여

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^k P(B_i)P(A|B_i)}$$

가 성립한다.



조건부 확률과 베이즈 정리

■ 베이즈 정리 예제

어떤 사람이 거짓말을 하고 있을 때 얼굴이 빨개질 확률이 0.6이고, 거짓말을 하지 않았는데도 얼굴이 빨개질 확률이 0.2라고 한다. 그리고 거짓말을 할 확률은 0.2라고 할 때, 이 사람이 이야기하는 도중에 얼굴이 빨개졌다면 지금 거짓말을 하고 있을 확률을 구하시오. 단, 얼굴이 빨개질 확률은 0.28 이라고 가정.

(풀이)

이 사람이 거짓말을 할 사상을 F , 얼굴이 빨개질 사상을 R 이라고 하면, $P(R|F) = 0.6, P(R|F^c) = 0.2, P(F) = 0.2, P(F^c) = 0.8$ 이다. $P(R) = 0.28$ 이고, 구하고자 하는 확률

$$P(F|R) \text{은 } P(F|R) = \frac{P(F \cap R)}{P(R)} = \frac{P(F)P(R|F)}{P(F)P(R|F) + P(F^c)P(R|F^c)} \text{ 이므로 } P(F|R) = \frac{0.2 \cdot 0.6}{0.28} = \frac{3}{7} \text{ 이다.}$$



조건부 확률과 베이즈 정리

■ 베이즈 정리 예제

[Example] The proportion of people in a given community who have a certain disease is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01.

If a person tests positive, what is the probability that the person actually has the disease?

[Solution] Let D represent the event that a person actually has the disease, and let $+$ represent the event that the test gives a positive signal. We wish to find $P(D|+)$. We know $P(D) = 0.005$, $P(+|D) = 0.99$, and $P(+|D^c) = 0.01$.

Using Bayes' rule:
$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.99(0.005)}{0.99(0.005) + 0.01(0.995)} = 0.332.$$



베이즈 결정 이론

통계적으로 최고의 분류기를 선택

다음의 내용들을 가정

- ✓ 결정 문제는 확률적인 문제로 표현이 가능
- ✓ 관련된 확률 정보들을 모두 알 수 있음 (현실에서는 알기 어려움)

관련용어

- ✓ 사전 확률 (prior probability)
- ✓ 사후 확률 (posterior probability)
- ✓ 우도 (likelihood)
- ✓ evidence

$$\begin{aligned} P(\omega_j | x) &= \frac{p(x | \omega_j) \cdot P(\omega_j)}{\sum_{x \in X} p(x | \omega_j) \cdot P(\omega_j)} \\ &= \frac{p(x | \omega_j) \cdot P(\omega_j)}{p(x)} \end{aligned}$$

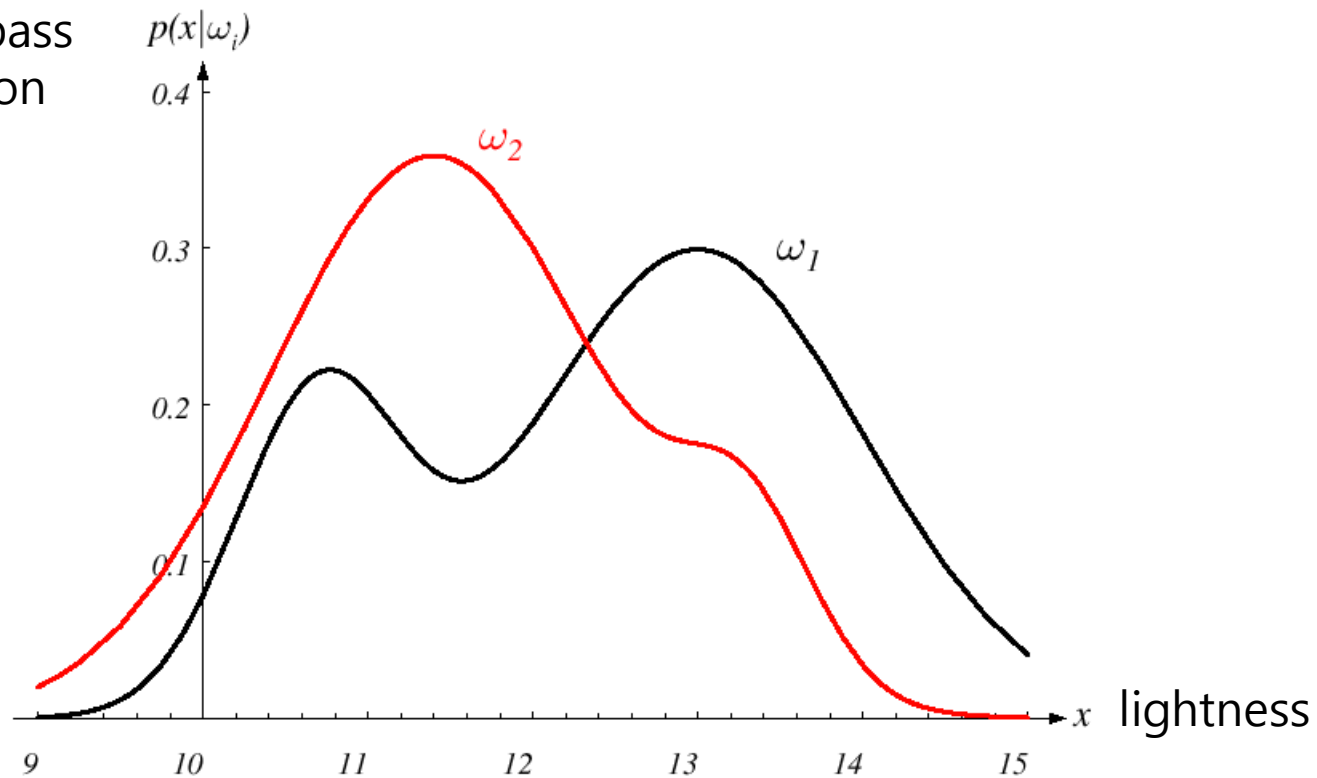


베이즈 결정 이론

$P(x|\omega_2)$: 클래스가 주어졌을 때 확률변수의 값

Likelihood: Salmon임이 관측되었을 때, lightness가 11 이상, 12 미만일 확률은 얼마인가?

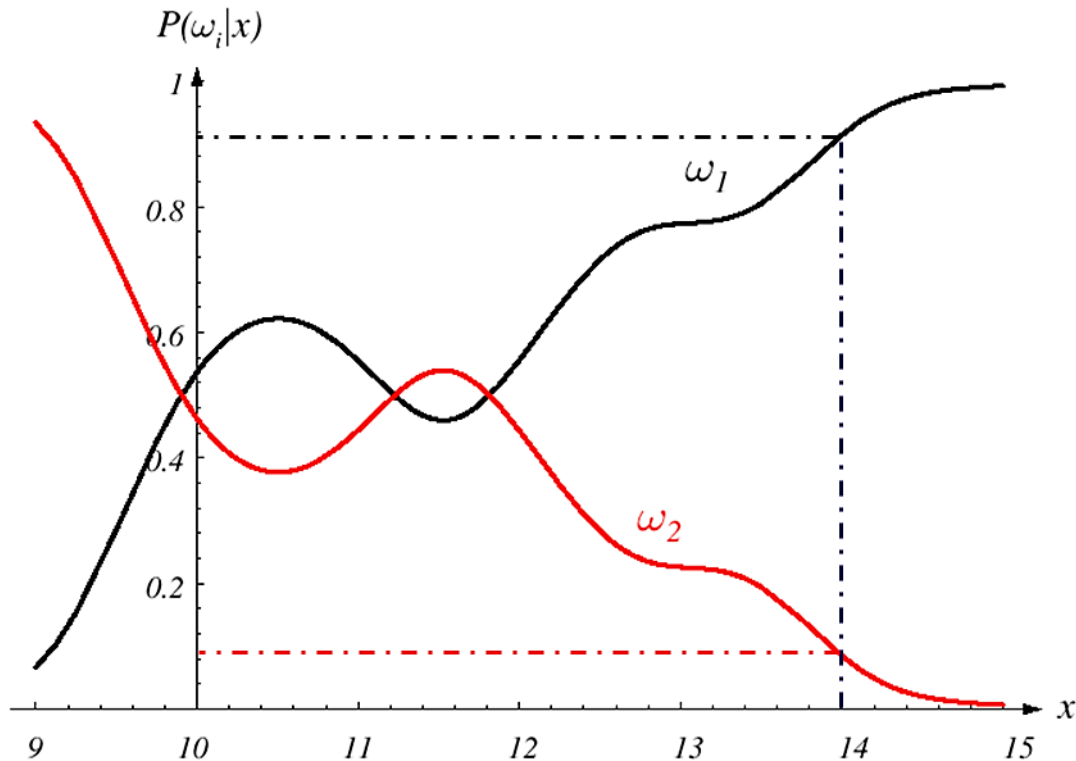
ω_1 : Sea bass
 ω_2 : Salmon





베이즈 결정 이론

■ 베이즈 정리를 사용하면 일반적으로 알기 어려운 사후 확률을 알기 쉬운 사전확률과 우도(그리고 evidence)를 이용하여 계산 가능



사전확률 $P(\omega_1) = 2/3$, $P(\omega_2) = 1/3$ 에 대한 사후확률 분포.

예를 들어 $x = 14$ 인 경우,

클래스 ω_2 에 속할 확률은 약 0.08 이고,

클래스 ω_1 에 속할 확률은 약 0.92 이다.

모든 x 에서, 사후 확률의 합은 1 이다.



베이즈 결정 이론

■ 베이즈 결정 이론: 사후 확률이 가장 높은 클래스로 분류

- ✓ 만약 모든 $i = 1, 2, \dots, c$ 에 대하여 $P(\omega_i|x) > P(\omega_j|x)$ 를 만족할 때,
 $P(error) = \min[P(\omega_1|x), P(\omega_2|x), \dots, P(\omega_c|x)]$ 에 해당하는 ω_i 를 선택한다.
- ✓ 다수의 특성 $X = \{x_1, x_2, \dots, x_d\}$ 이 있을 경우
모든 $i = 1, 2, \dots, c$ 에 대하여 $P(\omega_i|X) > P(\omega_j|X)$ 를 만족할 때,
 $P(error) = \min[P(\omega_1|X), P(\omega_2|X), \dots, P(\omega_c|X)]$ 에 해당하는 ω_i 를 선택한다.



단순 베이즈 분류 (Naïve Bayes Classifier)

■ 각 클래스 레이블을 random variable로 가정하면, input feature의 종류가 다양할 경우:

- ✓ 입력변수 (A_1, A_2, \dots, A_n)가 주어졌을 때,
 - ✓ 목표: 클래스 C 에 대한 예측
 - ✓ 구체적으로 다음 확률을 최대화하는 C 를 선정

$$P(C|A_1, A_2, \dots, A_n)$$

- ✓ 확률 $P(C|A_1, A_2, \dots, A_n)$ 를 데이터로부터 직접적으로 추정이 가능한가?



단순 베이즈 분류 (Naïve Bayes Classifier)

■ 각 클래스 레이블을 random variable로 가정하면, input feature의 종류가 다양할 경우:

✓ Approach:

- 베이즈 정리를 이용하여 모든 C에 대하여, 확률값 $P(C|A_1, A_2, \dots, A_n)$ 을 계산

$$P(C|A_1, A_2, \dots, A_n) = \frac{P(A_1, A_2, \dots, A_n|C)P(C)}{P(A_1, A_2, \dots, A_n)}$$

- 확률값 $P(C|A_1, A_2, \dots, A_n)$ 을 최대화하는 C를 선정
- $P(A_1, A_2, \dots, A_n|C)P(C)$ 를 최대화하는 것과 같은 의미
 - Why?

✓ $P(A_1, A_2, \dots, A_n|C)$ 는 어떻게 추정할 수 있을까 ?



단순 베이즈 분류 (Naïve Bayes Classifier)

■ 각 클래스 레이블을 random variable로 가정하면, input feature의 종류가 다양할 경우:

✓ 클래스가 주어졌을 때, 입력 변수 A_i 들 간의 독립을 가정하면:
(conditional independence)

- ✓ $P(A_1, A_2, \dots, A_n | C) = P(A_1 | C_j) P(A_2 | C_j) \dots P(A_n | C_j)$
- ✓ 모든 A_i 와 C_j 에 대하여 $P(A_i | C_j)$ 추정 가능
- ✓ 새로운 입력 데이터는 $P(C_j) \prod P(A_i | C_j)$ 를 최대화하는 C_j 로 분류



단순 베이즈 분류 (Naïve Bayes Classifier)

데이터로부터 확률의 추정

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

✓ Class: $P(C) = N_c/N$

- e.g., $P(No) = \frac{7}{10}$, $P(Yes) = \frac{3}{10}$

✓ For discrete attributes:

$$P(A_i|C_k) = |A_{ik}|/N_c$$

- $|A_{ik}|$ 는 클래스 C_k 에 속하는 데이터 중 입력변수 A_i 에 대항하는 데이터의 수
- Examples:

$$P(Status = Married|No) = 4/7$$

$$P(Refund = Yes|Yes) = 0$$



단순 베이즈 분류 (Naïve Bayes Classifier)

■ 데이터로부터 확률의 추정 (연속 값의 경우)

■ 연속 입력변수의 경우:

- ✓ 특정 범위로 이산화
 - ✓ 순서형 변수로 변환됨
 - ✓ 독립성 가정이 깨질 수 있음
- ✓ 이원분리: $(A < v) \text{ or } (A > v)$
 - ✓ 둘로 분리하여 새로운 변수 생성
- ✓ 확률밀도 추정
 - ✓ 입력변수들이 정규분포 등 특정 확률분포를 따르는 것으로 가정
 - ✓ 데이터를 이용하여 확률변수의 모수 추정
(e.g., mean and standard deviation)
 - ✓ 확률분포의 추정 이후, 이를 이용하여 $P(A_i|C)$ 계산



단순 베이즈 분류 (Naïve Bayes Classifier)

데이터로부터 확률의 추정 (연속 값의 경우)

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

✓ Normal distribution:

$$P(A_i|c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

- One for each (A_i, c_j) pair

✓ For (*Income*, *Class = No*):

- If Class=No
 - Sample mean = 110
 - Sample variance = 2975

$$P(\text{Income} = 120|\text{No}) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$



단순 베이즈 분류 (Naïve Bayes Classifier)

예시

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

naive Bayes Classifier:

$$P(\text{Refund}=\text{Yes}|\text{No}) = 3/7$$

$$P(\text{Refund}=\text{No}|\text{No}) = 4/7$$

$$P(\text{Refund}=\text{Yes}|\text{Yes}) = 0$$

$$P(\text{Refund}=\text{No}|\text{Yes}) = 1$$

$$P(\text{Marital Status}=\text{Single}|\text{No}) = 2/7$$

$$P(\text{Marital Status}=\text{Divorced}|\text{No}) = 1/7$$

$$P(\text{Marital Status}=\text{Married}|\text{No}) = 4/7$$

$$P(\text{Marital Status}=\text{Single}|\text{Yes}) = 2/7$$

$$P(\text{Marital Status}=\text{Divorced}|\text{Yes}) = 1/7$$

$$P(\text{Marital Status}=\text{Married}|\text{Yes}) = 0$$

For taxable income:

If class=No: sample mean=110
sample variance=2975

If class=Yes: sample mean=90
sample variance=25

- $$P(X|\text{Class}=\text{No}) = P(\text{Refund}=\text{No}|\text{Class}=\text{No}) \\ \times P(\text{Married}|\text{Class}=\text{No}) \\ \times P(\text{Income}=120\text{K}|\text{Class}=\text{No}) \\ = 4/7 \times 4/7 \times 0.0072 = 0.0024$$
- $$P(X|\text{Class}=\text{Yes}) = P(\text{Refund}=\text{No}|\text{Class}=\text{Yes}) \\ \times P(\text{Married}|\text{Class}=\text{Yes}) \\ \times P(\text{Income}=120\text{K}|\text{Class}=\text{Yes}) \\ = 1 \times 0 \times 1.2 \times 10^{-9} = 0$$

Since $P(X|\text{No})P(\text{No}) > P(X|\text{Yes})P(\text{Yes})$

Therefore $P(\text{No}|X) > P(\text{Yes}|X)$

=> Class = No



단순 베이즈 분류 (Naïve Bayes Classifier)

✓ 조건부 확률 중 하나 이상이 0인 경우 전체 확률이 0이 되는 문제가 발생

✓ 따라서 다음과 같은 확률 계산 방식을 통하여 이러한 문제를 해결

✓ 기본:
$$P(A_i | C) = \frac{N_{ic}}{N_c}$$

✓ 라플라스:
$$P(A_i | C) = \frac{N_{ic} + 1}{N_c + c}$$

✓ m-추정:
$$P(A_i | C) = \frac{N_{ic} + mp}{N_c + m}$$



단순 베이즈 분류 (Naïve Bayes Classifier)

● 정리

- ✓ 이상 값(outlier)에 강건함
- ✓ 결측 값을 조건부 확률 계산시 배제하여 자연스럽게 처리 가능
- ✓ 몇몇 attribute에 대하여는 독립성 가정이 성립하지 않을 수 있음
 - ✓ Bayesian Belief Network 등 다른 방법을 사용하여 해결



베이지안 추정 이론

- 도박사의 오류??
- 동전을 10번 던져서 앞면이 7번이 나왔다고 하자. 다음 번 던짐에서 앞면이 나올 확률은?
- 100번 던져서 앞면이 70번 나왔다면?
- 10만번 던져 앞면이 7만번 나왔다면?
- Frequentist vs Bayesian
- (주관적) 사전 정보의 반영



베이지안 추정 이론

✓ 베이지안 추정이론에서는 베이즈 확률 법칙을 이용하여 모수 θ 를 추정

✓ 사전 분포

$$P_{\Theta}(\theta)$$

▪ 사전 믿음을 반영

✓ 우도

$$P_{X|\Theta}(x|\theta)$$

▪ 모델을 표현

✓ 사후 분포

$$P_{\Theta|X}(\theta|D)$$

▪ 사전 분포와 데이터를 모두 반영한 사후 분포



베이지안 추정 이론

✓ 예측 분포

$$P_{X|T}(x | D) = \int P_{X|\Theta}(x | \theta) P_{\Theta|T}(\theta | D) d\theta$$

- 사후 확률을 이용한 새로운 데이터에 대한 예측 분포

✓ MAP (maximum a posteriori) 추정

- 예측 분포의 계산의 적분은 종종 계산이 어려움
- 이 때 다음과 같이 MAP를 이용하여 예측 분포를 근사

$$\begin{aligned}\theta_{MAP} &= \arg \max_{\theta} P_{\Theta|T}(\theta | D) \\ &= \arg \max_{\theta} P_{T|\Theta}(D | \theta) P_{\Theta}(\theta) \\ P_{X|T}(x | D) &= \int P_{X|\Theta}(x | \theta) \delta(\theta - \theta_{MAP}) d\theta \\ &= P_{X|\Theta}(x | \theta_{MAP})\end{aligned}$$



베이지안 선형 회귀 (optional)

■ 선형 회귀

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1x_1 + \dots + w_Dx_D$$

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x})$$

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$$

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1})$$

$$\mathbb{E}[t|\mathbf{x}] = \int t p(t|\mathbf{x}) dt = y(\mathbf{x}, \mathbf{w})$$



베이지안 선형 회귀 (optional)

■ 우도 (likelihood)

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1})$$

■ 로그-우도

$$\begin{aligned} \ln p(\mathbf{t}|\mathbf{w}, \beta) &= \sum_{n=1}^N \ln \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}) \\ &= \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D(\mathbf{w}) \end{aligned}$$

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2.$$



베이지안 선형 회귀 (optional)

■ 최대 우도 추정

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\} \phi(\mathbf{x}_n)^T.$$

$$0 = \sum_{n=1}^N t_n \phi(\mathbf{x}_n)^T - \mathbf{w}^T \left(\sum_{n=1}^N \phi(\mathbf{x}_n) \phi(\mathbf{x}_n)^T \right)$$

$$\mathbf{w}_{\text{ML}} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}$$



베이지안 선형 회귀 (optional)

■ (conjugate) 사전 분포

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)$$

■ 사후분포

$$p(\mathbf{w} | \mathbf{t}) = \mathcal{N}(\mathbf{w} | \mathbf{m}_N, \mathbf{S}_N)$$

$$\mathbf{m}_N = \mathbf{S}_N (\mathbf{S}_0^{-1} \mathbf{m}_0 + \beta \Phi^T \mathbf{t})$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta \Phi^T \Phi.$$

■ 등방향 사전 분포 가정하면

$$p(\mathbf{w} | \alpha) = \mathcal{N}(\mathbf{w} | \mathbf{0}, \alpha^{-1} \mathbf{I})$$

$$\mathbf{m}_N = \beta \mathbf{S}_N \Phi^T \mathbf{t}$$

$$\mathbf{S}_N^{-1} = \alpha \mathbf{I} + \beta \Phi^T \Phi.$$

$$\ln p(\mathbf{w} | \mathbf{t}) = -\frac{\beta}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 - \frac{\alpha}{2} \mathbf{w}^T \mathbf{w} + \text{const.}$$



베이지안 선형 회귀 (optional)

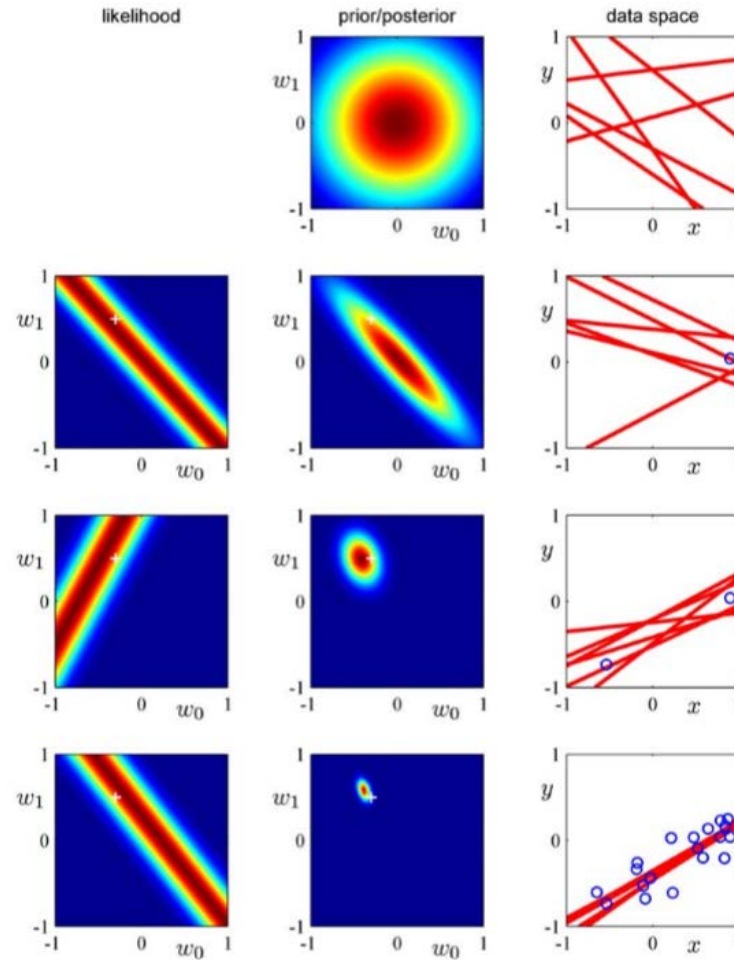


Figure 3.7 Illustration of sequential Bayesian learning for a simple linear model of the form $y(x, \mathbf{w}) = w_0 + w_1 x$. A detailed description of this figure is given in the text.



베이지안 선형 회귀 (optional)

예측 분포

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) d\mathbf{w}$$

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m}_N^T \phi(\mathbf{x}), \sigma_N^2(\mathbf{x}))$$

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \phi(\mathbf{x})^T \mathbf{S}_N \phi(\mathbf{x})$$

When full Bayesian??

$$p(t|\mathbf{t}) = \iiint p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) p(\alpha, \beta|\mathbf{t}) d\mathbf{w} d\alpha d\beta$$

MAP 추정??



베이지안 선형 회귀 (optional)

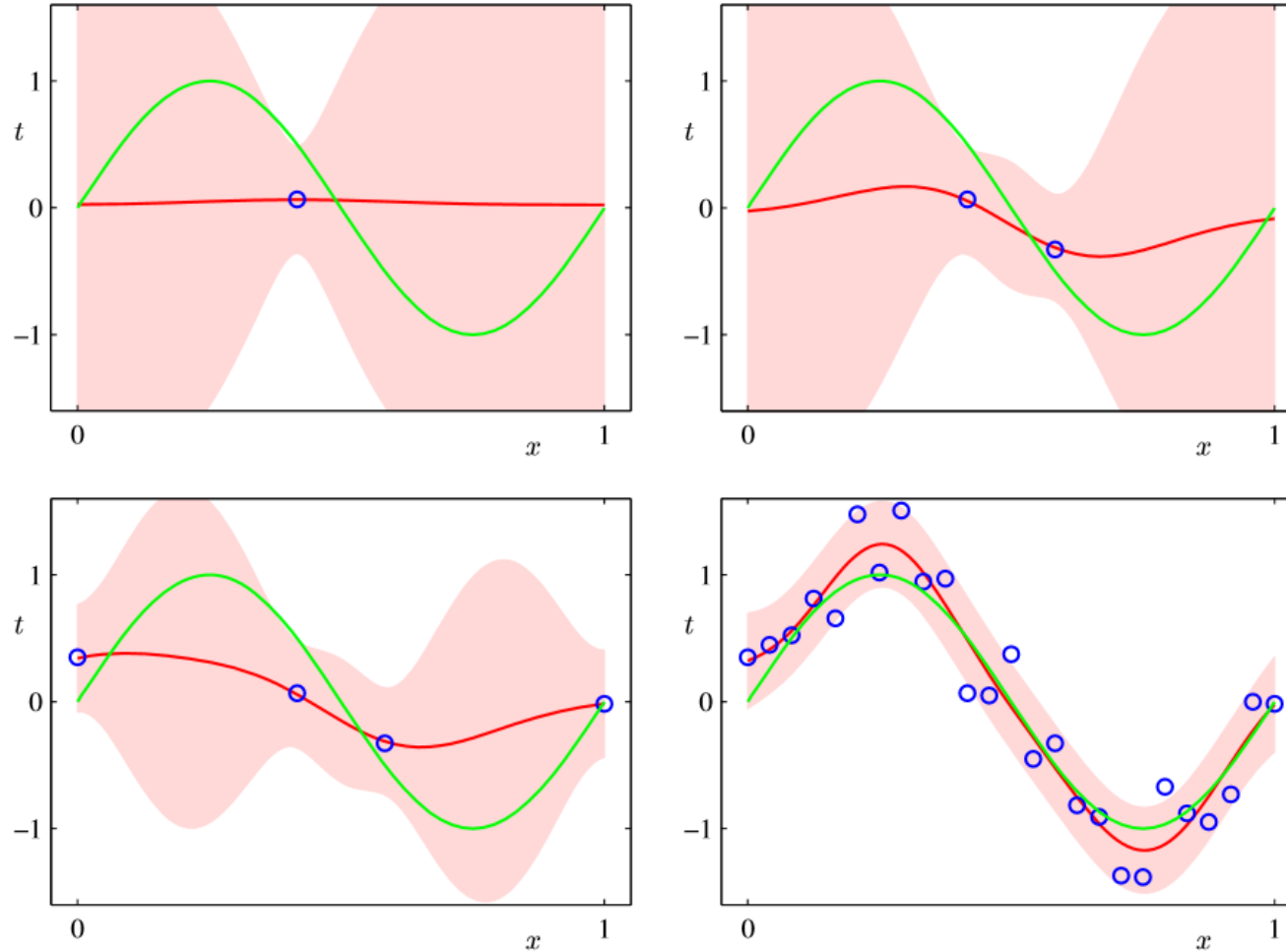


Figure 3.8 Examples of the predictive distribution (3.58) for a model consisting of 9 Gaussian basis functions of the form (3.4) using the synthetic sinusoidal data set of Section 1.1. See the text for a detailed discussion.



베이지안 선형 회귀 (optional)

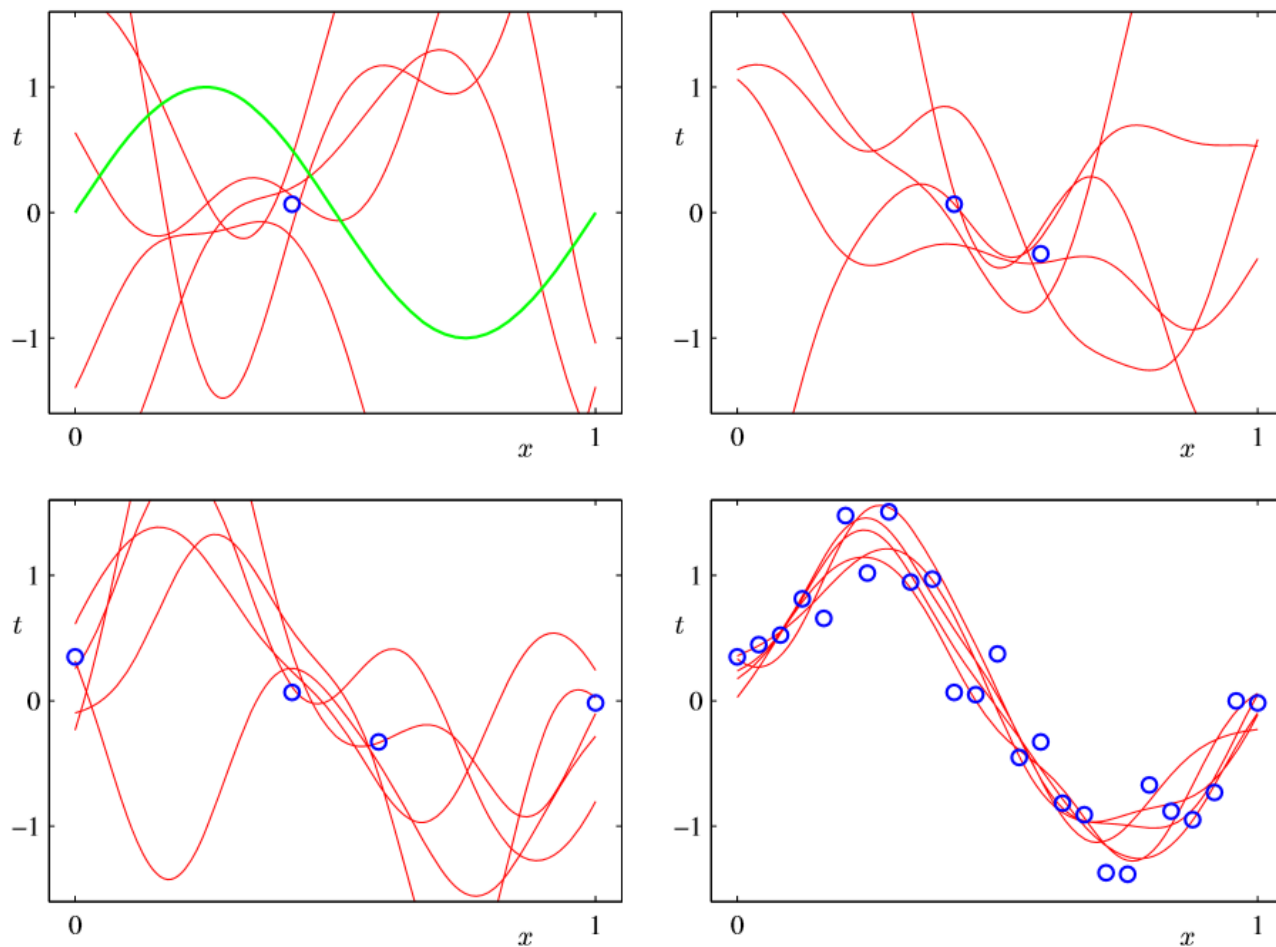


Figure 3.9 Plots of the function $y(x, w)$ using samples from the posterior distributions over w corresponding to the plots in Figure 3.8.



베이지안 선형 회귀 (optional)

MAP

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) d\mathbf{w}$$

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(t|\mathbf{m}_N^T \phi(\mathbf{x}), \sigma_N^2(\mathbf{x}))$$

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \phi(\mathbf{x})^T \mathbf{S}_N \phi(\mathbf{x})$$

When full Bayesian??

$$p(t|\mathbf{t}) = \iiint p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta) p(\alpha, \beta|\mathbf{t}) d\mathbf{w} d\alpha d\beta$$



실습 - 모듈 설치 및 import

Seaborn 패키지 설치(visualization 패키지)

```
(machinelearning) C:\Users\seokwon>conda install seaborn
```

필요 패키지 import

```
from sklearn.datasets import load_iris
import pandas as pd
import numpy as np
```



실습 - Iris Data load

```
iris = load_iris()
iris_frame=pd.DataFrame(data=np.c_[iris['data'],iris['target']],columns=iris['feature_names'] + ['target'])
iris_frame['target'] = iris_frame['target'].map({0:"setosa",1:"versicolor",2:"virginica"})
X=iris_frame.iloc[:, :-1]
Y=iris_frame.iloc[:, [-1]]
iris_frame
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica



실습 - Iris Data load

Iris Dataset

✓ 데이터 수 : 150

◆ Feature

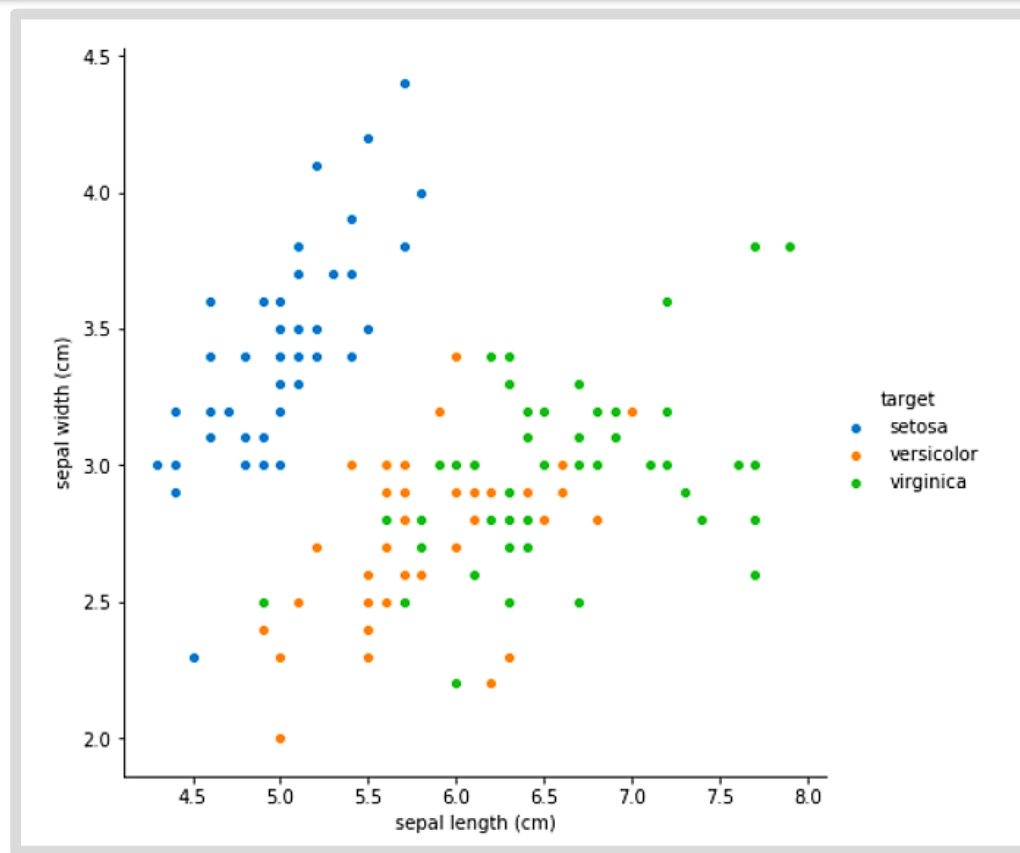
- ✓ Sepal Length : 꽃받침의 길이
- ✓ Sepal Width : 꽃받침의 너비
- ✓ Petal Length : 꽃잎의 길이
- ✓ Petal Width : 꽃잎의 너비
- ✓ Species: 꽃의 종류, setosa/versicolor/virginina(Target)



실습 - Iris Data visualization

```
import seaborn as sn
```

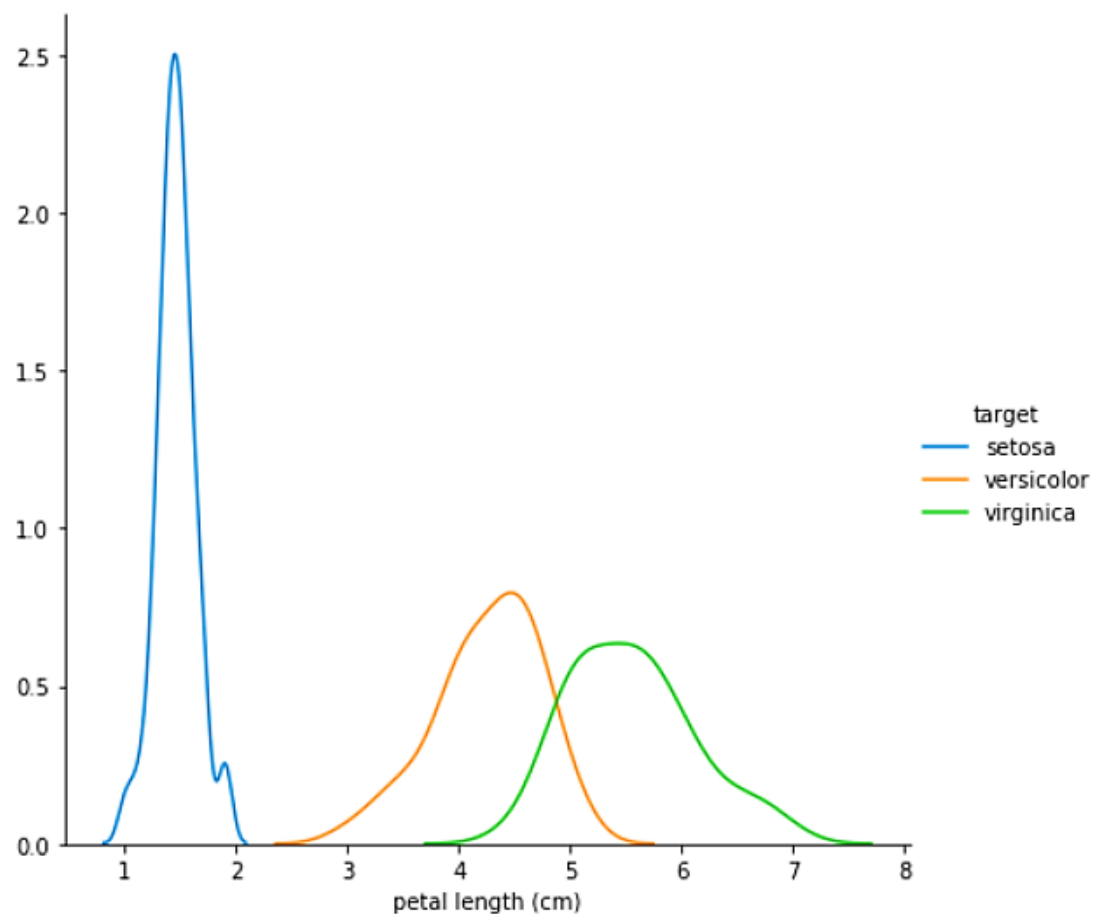
```
sn.pairplot(iris_frame, x_vars=["sepal length (cm)"], y_vars=["sepal width (cm)"], hue="target", size=3)
```





실습 - Iris Data visualization

```
sn.FacetGrid(iris_frame, hue="target", size=6).map(sn.kdeplot, "petal length (cm)").add_legend()
```

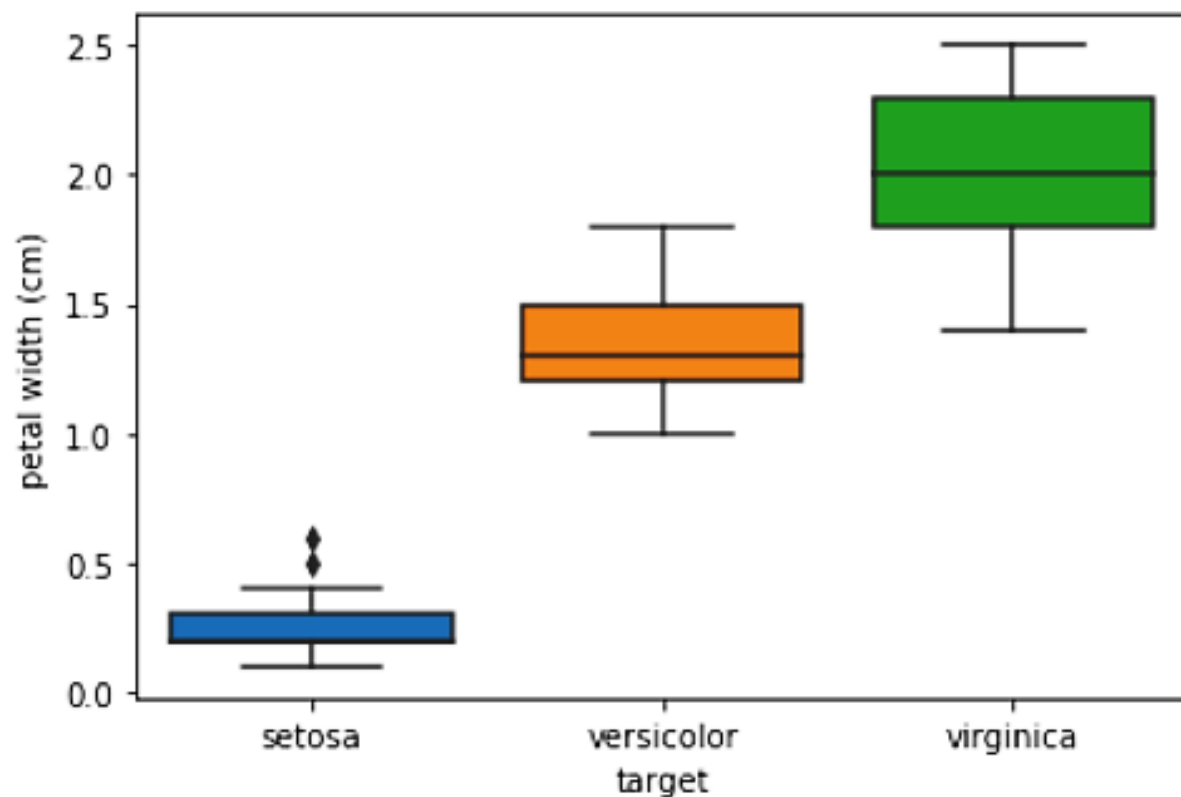




실습 - Iris Data visualization

```
sn.bboxplot(x="target",y="petal width (cm)", data=iris_frame)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x2091a67ef10>
```





실습 - 모델 학습

■ Feature 중 sepal에 관련된 두 개의 feature만 이용해서 학습

```
import matplotlib.colors as colors
df1 = iris_frame[["sepal length (cm)", "sepal width (cm)", 'target']]
X = df1.iloc[:,0:2]
Y = df1.iloc[:,2].replace({'setosa':0, 'versicolor':1, 'virginica':2}).copy()
NB.fit(X,Y)
N = 100
```




실습 - 모델 학습 시각화

```
X_ = np.linspace(4, 8, N)
Y_ = np.linspace(1.5, 5, N)
X_, Y_ = np.meshgrid(X_, Y_)

color_list = ['Blues', 'Greens', 'Reds']
my_norm = colors.Normalize(vmin=-1., vmax=1.)
g = sns.FacetGrid(iris_frame, hue="target", size=10, palette = 'colorblind') .map(plt.scatter, "sepal length (cm)",
                                                                                   "sepal width (cm)",).add_legend()

my_ax = g.ax
zz = np.array( [NB.predict( [[xx,yy]])[0] for xx, yy in zip(np.ravel(X_), np.ravel(Y_)) ] )
Z = zz.reshape(X_.shape)
```

■ 격자 안의 모든 점을 예측하고 해당 예측을 통해서 decision boundary visualization

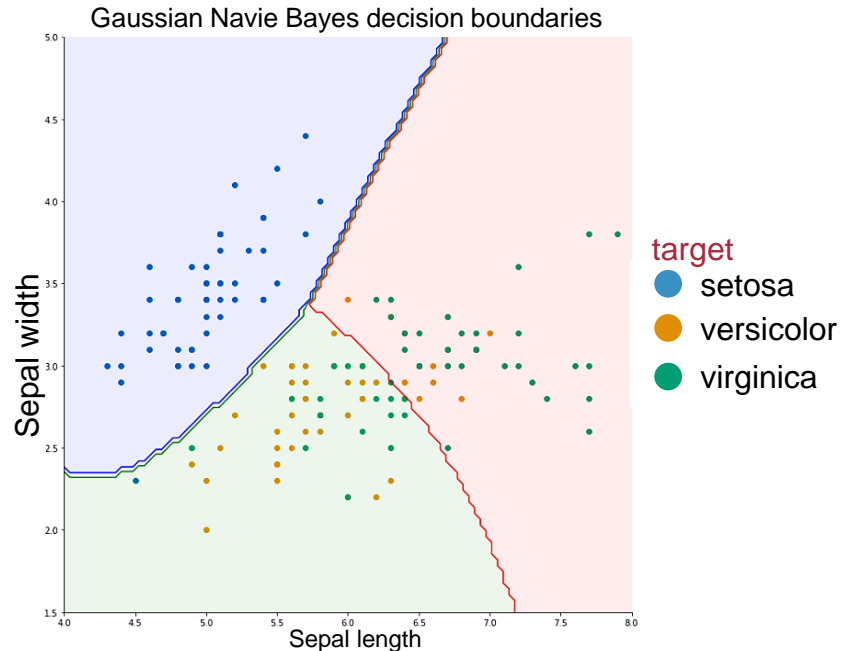


실습 - 모델 학습

```
#Plot the filled and boundary contours
my_ax.contourf( X_, Y_, Z, 2, alpha = .1, colors = ('blue','green','red'))
my_ax.contour( X_, Y_, Z, 2, alpha = 1, colors = ('blue','green','red'))

# Addd axis and title
my_ax.set_xlabel('Sepal length')
my_ax.set_ylabel('Sepal width')
my_ax.set_title('Gaussian Naive Bayes decision boundaries')

plt.show()
```





실습 - 모델 학습

