

13강. SVM과 커널법

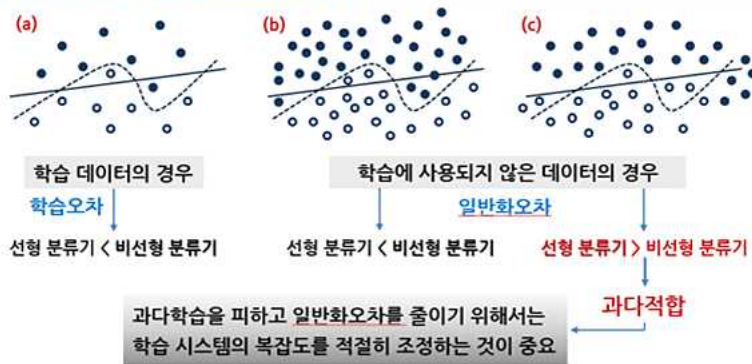
※ 점검하기

Q1. 선형 분류기의 장단점과 커널법의 필요성에 대해 설명하시오.

〈관련학습보기〉

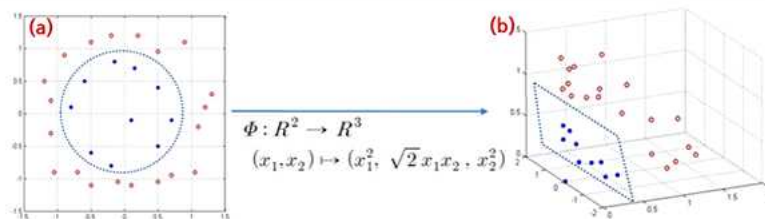
1) 과다적합과 일반화오차

학습 시스템의 복잡도와 일반화오차의 관계



1) 커널의 필요성

저차원의 입력 x 를 고차원의 공간의 값 $\phi(x)$ 로 매핑시키는 함수 ϕ



고차원 문제로 선형화하면 간단한 선형분류기를 사용한 분류가 가능

계산량 증가와 같은 부작용 발생 → 「커널법」으로 해결

고차원 매핑을 통해 비선형 문제를 선형화하여 해결하면서
커널 함수를 통해 계산량 증가의 문제를 해결하는 방법

선형 분류기는 계산 복잡도가 간단하지만 데이터 분포가 비선형일 경우에 적합하지 않다.

이런 관점에서 계속해서 설명을 해 보세요.

[참조] 1. 선형 초평면에 의한 분류의 「1) 과다적합과 일반화오차」

3. 커널법의 「1) 커널의 필요성」

Q2. 다층 퍼셉트론에 의한 분류기와 SVM에 의한 분류기에 대해서 다음 사항을 중심으로 특성을 비교해 보시오.

(1) 결정경계의 형태

(2) 학습의 목적

(3) 학습과 인식에 필요한 계산량

(4) 학습을 통해 찾아지는 최적해의 특성

<관련학습보기>

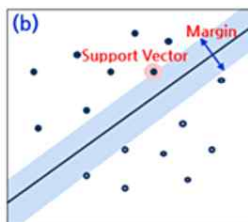
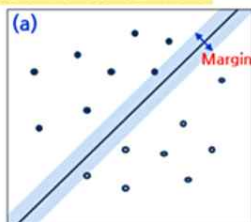
1) SVM(Support Vector Machine)

1 『최대 마진 분류기(Maximum Margin Classifier)』

» **마진** → 학습데이터들 중에서 결정경계에 가장 가까운 데이터로부터 결정경계까지의 거리

» **서포트벡터** → 결정경계에 가장 가까운 곳에 위치한 데이터

결정경계에 따른 마진의 차이



일반화오차를 작게
↓
클래스 간의 간격을 최대로
↓
마진을 최대로
↓
『최대 마진 분류기』, 『SVM』

2) 커널법과 SVM

n 차원의 입력 x 를 m 차원의 특징데이터 $\phi(x)$ 로 변환한 후 SVM을 이용해서 분류한다고 가정

SVM에서 연산은 개개의 $\phi(x)$ 가 아니라 두 벡터의 내적 $\phi(x) \cdot \phi(y)$ 를 사용

고차원 매핑 $\phi(x)$ 를 정의하는 대신에 $\phi(x) \cdot \phi(y)$ 를 하나의 함수 $k(x, y)$ 로 정의하여 사용

$$\begin{aligned} k(x, y) &= \phi(x) \cdot \phi(y) \\ &= (x_1^2, \sqrt{2}x_1x_2, x_2^2) \cdot (y_1^2, \sqrt{2}y_1y_2, y_2^2) \\ &= (x \cdot y)^2 \end{aligned}$$

커널 함수

파라미터 추정을 위한 라그랑주 함수 $L(w, w_0, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \{y_i(w^T \phi(x_i) + w_0) - 1\}$

이원적 문제의 함수 $Q(\alpha)$
$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \phi(x_i) \cdot \phi(x_j)$$

$$= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

2) 커널법과 SVM

커널 함수만으로 표현된 분류 함수

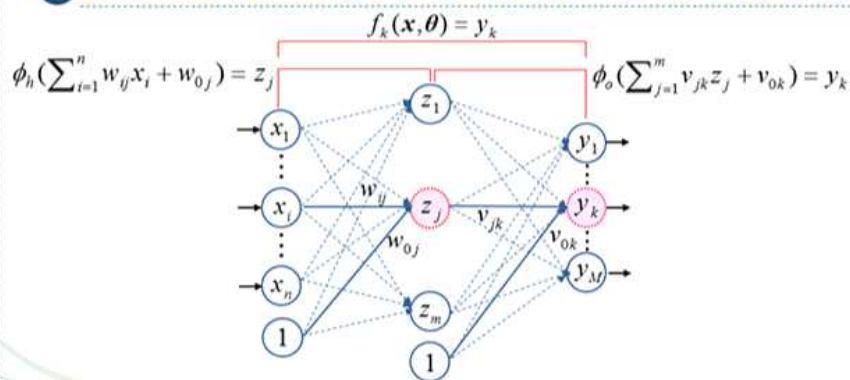
$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i \phi(x) \cdot \phi(x_i) + w_0 \right) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i k(x, x_i) + w_0 \right)$$

대표적인 커널 함수

| | |
|----------|--|
| 선형 커널 | $k(x, y) = (x \cdot y)$ |
| 다항식 커널 | $k(x, y) = (x \cdot y + c)^d$ |
| 시그모이드 커널 | $k(x, y) = \tanh(\theta_1 x \cdot y + \theta_2)$ |
| 가우시안 커널 | $k(x, y) = \exp \left\{ -\frac{\ x - y\ ^2}{2\sigma^2} \right\}$ |

3) 다층 퍼셉트론

1 MLP(Multilayer Perceptron), 1개 이상의 은닉층을 가진 다층 전방향 신경망



다층 퍼셉트론과 선형 SVM 및 커널 SVM을 비교해보세요.

[참조] 13강. SVM과 커널법 - 「2. 서포트벡터머신」, 「3. 커널법」

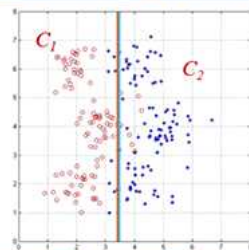
12강. 신경망 - 「2. 신경망 분류기」

Q3. 슬랙변수가 필요한 이유에 대해 설명하시오.

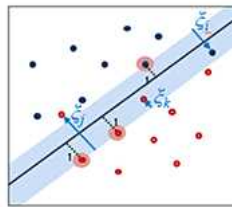
<관련학습보기>

4) 슬랙변수를 가진 SVM

선형 분리가 불가능한 데이터를 위해 슬랙변수 ξ 도입



잘못 분류된 데이터로부터 해당 클래스의 결정경계까지의 거리



슬랙변수를 포함한 분류 조건

$$\begin{cases} (w^T x_i + w_0) \geq +1 - \xi_i & \text{for } y_i = +1 \\ (w^T x_i + w_0) \leq -1 + \xi_i & \text{for } y_i = -1 \end{cases} \rightarrow y_i (w^T x_i + w_0) \geq 1 - \xi_i \quad (i = 1, \dots, N)$$

슬랙변수의 값이 클수록 더 심한 오분류를 허용!

교재 12.3.1항(슬랙변수의 도입)을 참고하세요.

[참조] 2. 서포트벡터머신의 「4) 슬랙변수를 가진 SVM」

※ 정리하기

1. 선형 초평면에 의한 분류

- 1) 일반화오차를 작게 하기 위해서는 두 클래스 간의 간격을 최대로 하는 것이 좋고, 따라서 마진을 최대로 하는 결정경계를 찾는 것이 바람직함
- 2) 이와 같은 목적에 맞춰 최적화된 선형 결정경계를 찾는 분류기를 최대 마진 분류기, 일반적으로 서포트벡터머신(SVM)이라고 함

2. 서포트벡터머신

- 1) 슬랙변수란?
 - 슬랙변수는 데이터가 해당 클래스의 결정경계를 넘어서 다른 클래스 영역에 존재할 수 있도록 허용함
 - 따라서 슬랙변수가 클수록 더 심한 오분류를 허용함을 의미

3. 커널법

- 1) 커널법이란?
 - : 입력 데이터의 차원을 높임으로써 문제를 선형화하면 간단한 선형 분류기를 사용하여 분류할 수 있지만, 차원의 증가로 인한 계산량 증가와 같은 부작용도 발생하는데 이를 위해 제안된 것이 커널법이라고 함
- 2) 커널 함수
 - : SVM의 연산 과정에서는 고차원 매핑 $\phi(\mathbf{x})$ 의 값이 아니라 두 벡터의 내적을 사용하므로, $\phi(\mathbf{x})$ 대신에 $\phi(\mathbf{x}) \cdot \phi(\mathbf{y})$ 를 하나의 함수 $k(\mathbf{x}, \mathbf{y})$ 로 정의하여 사용할 수 있으며, 이 함수를 커널 함수라고 함