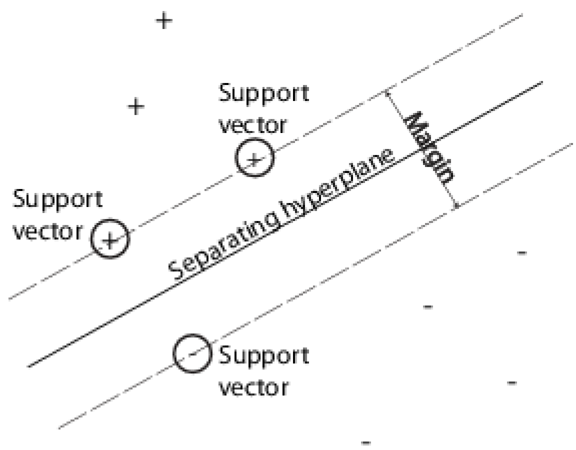
CHPATER5. 서포트 벡터 머신

- 선형 SVM 분류
- 비선형 SVM 분류
- SVM 회귀
- SVM 이론

서포트 벡터 머신 (Support Vector Machine, SVM)

- SVM은 매우 강력하고 선형이나 비선형 분류, 회귀, 이상치 탐색에도 사용할 수 있습니다.
- SVM은 특성 스케일이 민감합니다.
- SVM은 특히 복잡한 분류 문제에 잘 들어맞으며 작거나 중간 크기의 데이터셋에 적합합니다.

5.1 선형 SVM 분류



- 그림에서 실선은 SVM 분류기의 결정 경계입니다.
- 이 직선은 두 개의 클래스를 나눌뿐 아니라 제일 가까운 훈련 샘플로부터 가능한 멀리 떨어져있습니다.
- SVM은 두 분류의 사이의 여백(Margin)을 최대화 하여 분류 능력을 극대화하는 선형분류기입니다. (라지 마진 분류)
- 도로 경계에 위치한 샘플을 서포트 벡터 (Support Vector)라 합니다.

5.1.1 하드 마진 분류 (Hard Margin Classification)

- 모든 샘플들이 도로 바깥쪽이 올바르게 분류되어 있는 것을 의미합니다.
- 하드 마진 분류는 데이터가 선형적으로 구분될 수 있어야 제대로 작동하고 이상치에 민감한 단점이 있습니다.

5.1.2 소프트 마진 분류 (Soft Margin Classification)

- 하드 마진 분류의 단점을 보완하기 위한 방법입니다.
- 도로의 폭을 가능한 넓게 유지하는 것과 마진 오류 사이에 적절한 균형을 잡는 방법입니다.
- Scikit-learn의 SVM 모델에선 C 하이퍼파라미터를 이용해 Margin을 조절할 수 있씁니다.
- C를 줄이면 Margin이 넓어집니다. 일반화 가능성이 높습니다.
- C를 높이면 Margin이 좁아집니다. 일반화 가능성이 낮습니다. (오버피팅 가능성)

5.2 비선형 SVM 분류

5.2.1 다항식 커널

- 낮은 차수의 다항식은 매우 복잡한 데이터 셋을 잘 표현하지 못하고 높은 차수의 다항식은 많은 특성을 추가하므로 모델을 느리게 만듭니다.
- SVM을 사용할 땐 커널 트릭 (Kernel trick)을 적용할 수 있습니다.
- 커널트릭은 실제로 특성을 추가하지 않으면서 다항식의 특성을 많이 추가한것과 같은 결과를 얻을 수 있습니다.

5.2.2 유사도 특성 추가

- 각 샘플이 특정 랜드마크와 얼마나 닮았는지 측정하는 유사도 함수로 계산한 특성을 추가하는 것입니다.
- 가우시안 방사 기저함수 등을 유사도 함수로 정의해 데이터를 재정의 합니다.

5.2.3 가우시안 RBF 커널

$$\phi_\gamma(x, l) = exp(-\gamma \| x - l \|^2)$$

- γ를 증가시키면 종 모양 그래프가 좁아져서 각 샘플의 영향 범위가 작아집니다. 그리고 결정 경계가 구불해집니다.
- γ를 감소시키면 종 모양 그래프가 넓어져서 각 샘플의 영향 범위가 커집니다. 그리고 결정 경계가 부드러워집니다.
- 오버피팅일땐 γ를 감소시키고 언더피팅일땐 γ를 증가시킵니다.

5.2.4 계산 복잡도

클래스	시간 복잡도	외부메모리 학습지원	스케일조정의 필요성	커널트릭
LinearSVC	O(m x n)	NO	YES	NO
SGDClassifier	O(m x n)	YES	YES	NO
SVC	O(m^2 x n) ~ O(m^3 x n)	NO	YES	YES

5.3 SVM 회귀

- SVM 회귀를 사용하는 방법은 분류와 목표를 반대로 하면 됩니다.
- 제한된 마진 오류 안에서 도로 안에 가능한 많은 샘플이 들어가도록 학습합니다.
- 도로의 폭은 하이퍼파라미터 ϵ으로 조절합니다.

5.4 SVM 이론

5.4.1 결정함수와 예측

$$w^T \cdot x + b < 0 \text{ 일때 } \hat{y} = 0$$
$$w^T \cdot x + b \geq 0 \text{ 일때 } \hat{y} = 1$$

- 선형 SVM 분류기 모델은 단순히 결정함수 $w^T \cdot x + b = w_1 x_1 + \cdots + w_n x_n + b$를 계산해 예측합니다.
- 선형 SVM 분류기를 훈련하는 것은 가능한 마진을 크게 하는 w와 b를 찾는 것을 의미합니다.

5.4.2 목적 함수

1) 하드 마진 선형 SVM 분류기의 목적함수

$$minimize \ \frac{1}{2} w^T \cdot w$$

[조건] $i = 1, 2, \ldots, m$ 일때 $t^{(i)}(W^T \cdot x^{(i)} + b) \geq 1$

- 가중치 벡터 w가 작을수록 마진이 커집니다.

2) 소프트 마진 선형 SVM 분류기의 목적함수

$$minimize \ \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} \varsigma^i$$

[조건] $i = 1, 2, \ldots, m$ 일때 $t^{(i)}(W^T \cdot x^{(i)} + b) \geq 1 - \varsigma^{(i)}$ 이고 $\varsigma^{(i) \geq 0}$

5.4.3 콰드라틱 프로그래밍

- 하드 마진과 소프트 마진 문제와 같이 선형적인 제약 조건이 있는 볼록함수의 이차 최적화 문제를 의미합니다.

$$minimize \ \frac{1}{2} p^T \cdot H \cdot p + f^T \cdot p$$

[조건] $A \cdot p \leq b$

- p는 n_p 차원의벡터를 의미합니다. (n_p = 모델 파라미터 수)
- H는 $n_p x n_p$ 크기의 행렬을 의미합니다.
- f는 n_p 차원의 벡터를 의미합니다.
- A는 $n_c x n_p$ 크기의 행렬을 의미합니다. (n_c = 제약수)
- b는 n_c 차원의 벡터를 의미합니다.

5.4.4 쌍대문제

- 원 문제 (Primal Problem)라는 제약이 있는 최적화 문제가 주어진다면 쌍대 문제 (Dual Problem)라고 하는 깊게 관련된 다른 문제로 표현할 수 있습니다.
- 훈련 샘플 수가 특성 개수보다 작을 때 원 문제보다 쌍대 문제를 푸는 것이 더 빠릅니다.

5.4.5 커널 SVM

- 머신러닝에서 커널은 변환 ϕ를 계산하지 않고 원래 벡터 a와 b에 기반하여 점곱을 계산할 수 있는 함수입니다.

[대표적인 커널]

- 선형 : $K(a, b) = a^T \cdot b$
- 다항식 : $K(a, b) = (\gamma a^T \cdot b + r)^d$
- 가우시안 RBF : $K(a, b) = exp(\gamma \|a - b\|^2)$
- 시그모이드 : $K(a, b) = tanh(\gamma a^T \cdot b + r)$

5.4.6 온라인 SVM

- 선형 SVM 분류기에 사용할 수 있는 한 가지 방법은 원 문제로부터 유도된 비용 함수를 최소화하기 위한 경사하강법을 사용하는 겁니다.
- 하지만 이 방식은 QP 기반의 방법보다 훨씬 느리게 수렴합니다.

$$J(w, b) = \frac{1}{2} w^T \cdot w + C \sum_{i=1}^{m} max(0, 1 - t^{(i)}(w^T \cdot x^{(i)} + b))$$

- 첫 번째 항은 모델이 작은 가중치 벡터 w를 가지도록 제약을 가해 마진을 크게 만듭니다.
- 두 번째 항은 모든 마진 오류를 계산합니다. 이 항을 최소화하면 마진오류를 가능한 한 줄이고 크기도 작게 만듭니다.