LinearSVC, # SVM이 아니라 왜 SVC일까?

SVM (Support Vector Machine)

- **LinearSVC :** Linear Support Vector Classification
- **LinearSVR :** Linear Support Vector Regression
- **OneClassSVM :** Unsupervised Outlier Detection
- **SVC :** C-Support Vector Classification
- **SVR :** Epsilon-Support Support Vector Regression

C가 붙는 이유는 Classification이라서 그렇다.

조성진

선형 SVM 분류 여러방법의 sklearn 코드

StandardScaler + LinearSVC(C=1, loss="hinge")

SVC(kernel="linear", C=1)

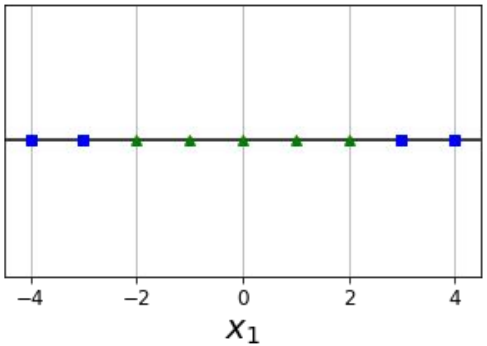
SGDClassifier(loss="hinge", alpha=1/(m*C)) : 일반적 SGD 사용 (온라인 학습의 장점 사용)

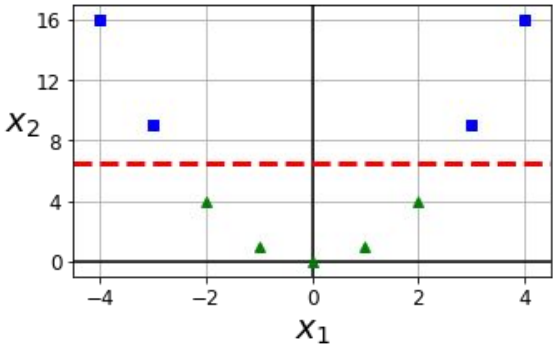
비선형 데이터셋에 대처하는 방법

Page 209

그림에서 볼 수 있듯이 이 데이터셋은 선형적으로 구분이 안 됩니다. 하지만 두 번째 특성을 추가하여 만들어진 2차원 데이터셋은 완벽하게 선형적으로 구분할 수 있습니다.

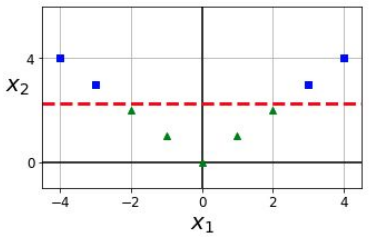
```
X1D = np.linspace(-4, 4, 9).reshape(-1, 1)
X2D = np.c_[X1D, X1D**2]
```



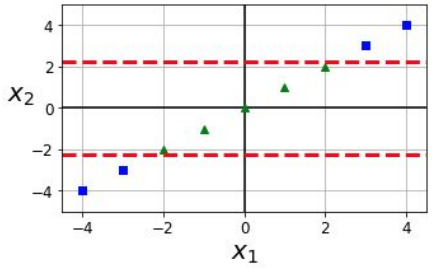


Quiz) accuray score가 1이 나오는 데이터는?

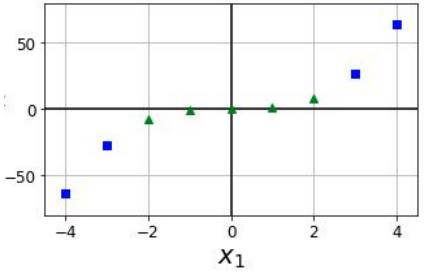
```
X1D = np.linspace(-4, 4, 9).reshape(-1, 1)
X2D = np.c_[X1D, abs(X1D)]
```




```
X1D = np.linspace(-4, 4, 9).reshape(-1, 1)
X2D = np.c_[X1D, X1D]
```



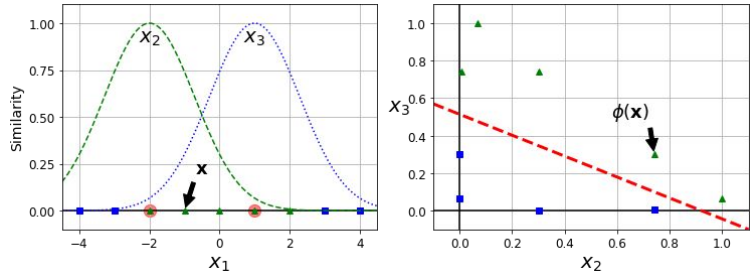

```
X1D = np.linspace(-4, 4, 9).reshape(-1, 1)
X2D = np.c_[X1D, X1D**3]
```



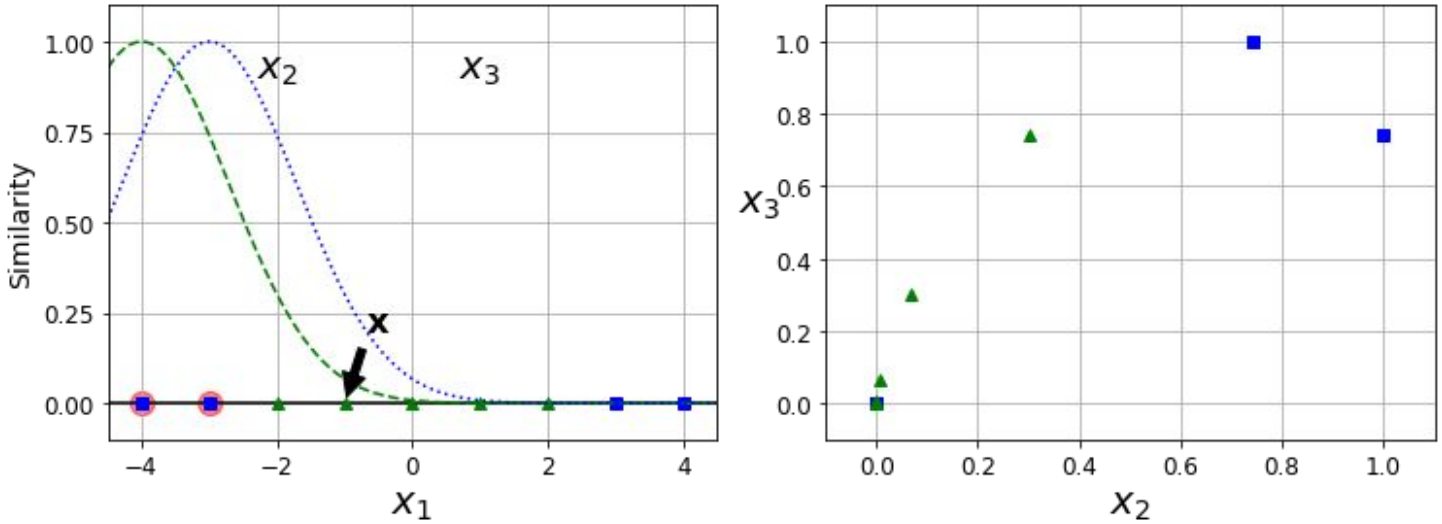

```
from sklearn.svm import LinearSVC
from sklearn.metrics import accuracy_score

ls = LinearSVC(C=1, loss="hinge")
ls.fit(X2D, y)
pred = ls.predict(X2D)
print(accuracy_score(pred, y))
```

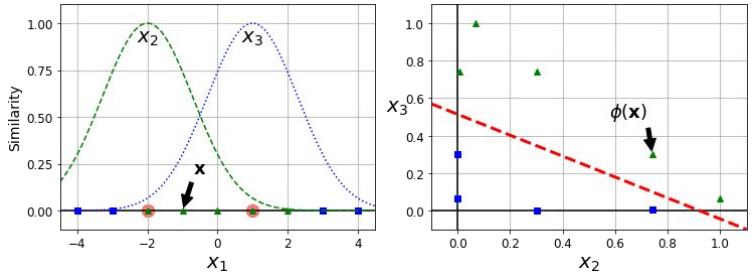
P-212, 유사도 함수와 랜드 마크



적절한 랜드마크를 설정해야지만
유의미한 특성을 추가할 수 있을까?



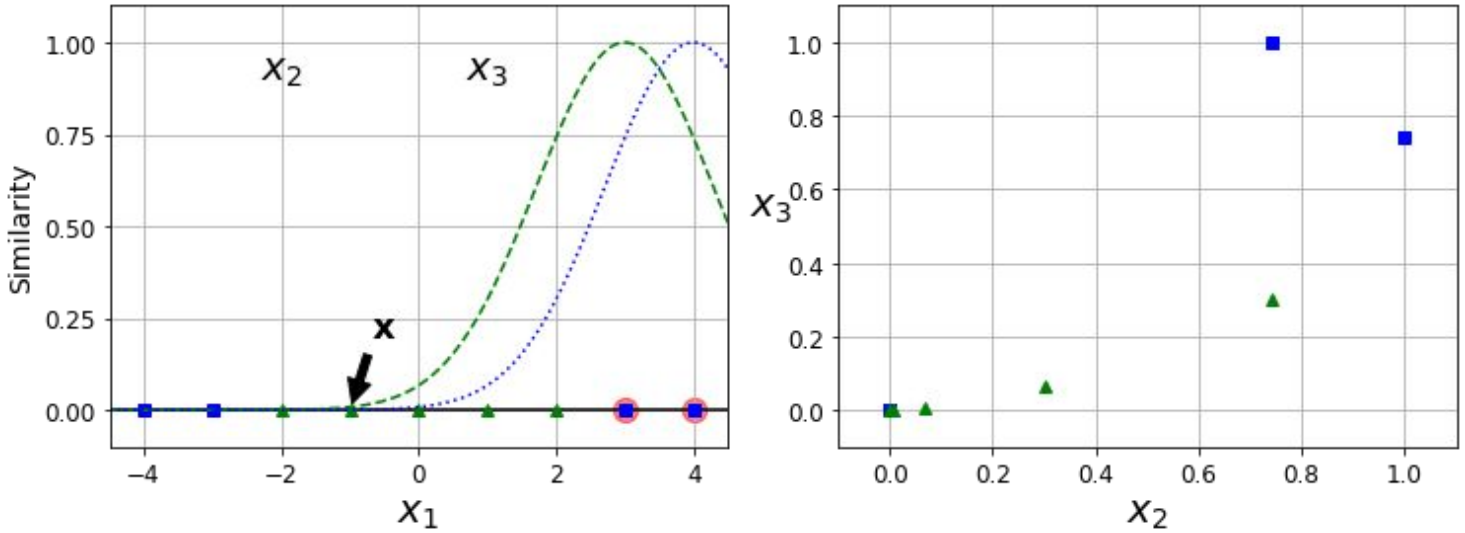
P-212, 유사도 함수와 랜드 마크



적절한 랜드마크를 설정해야지만

유의미한 특성을 추가할 수 있을까? **그렇다.**

=> 교재에서는 모든 샘플에 랜드마크



설명

김유지

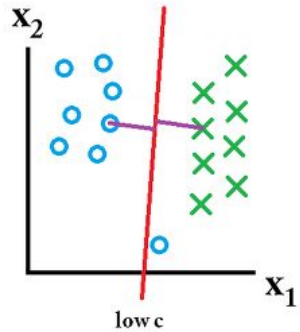
SVM 함수에서 hyper-parameter C는 무엇인가?

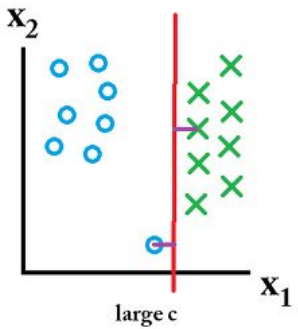
p. 207

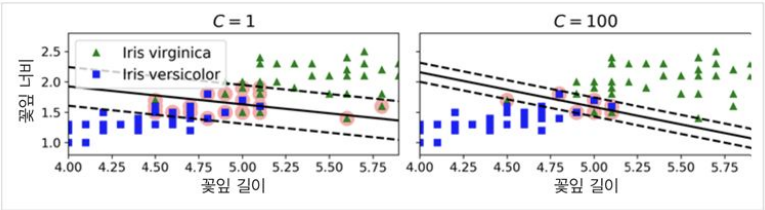
마진 오류는 샘플이 도로 중간이나 반대쪽에 있는 경우
마진오류는 나쁘므로 일반적으로 적은 것이 좋지만 마진 오류가 많은 왼쪽 모델이 일반화가 잘 될 것 같다
SVM 모델이 과대적합이면 C 감소시켜 모델 규제 가능하다

C는 cost로, **얼마나 많은 데이터 샘플이 다른 클래스에 놓이는 것을 허용하는지 결정**
C값을 낮게 설정 -> 이상치 있을 가능성 크게 잡아 일반적 결정 경계 찾음
C값을 높게 설정 -> 이상치의 존재 가능성 작게 봐 세심한 결정 경계 찾는다

C 너무 낮으면 과소적합, C 너무 높으면 과대적합 될 수 있다
C의 유무에 따라 하드마진, 소프트마진 SVM이라고 불린다







https://bskyvision.com/163

질문

백서윤

설명

구건모

원 문제(primal problem)와 쌍대 문제(dual problem)

P-223

원 문제라는 제약이 있는 최적화 문제가 주어지면
쌍대 문제라고 하는 깊게 관련된 다른 문제로 표현할 수 있습니다.

일반적으로 쌍대 문제 해는 원 문제 해의 하한값이지만, …

> 원 문제가 뭐고 쌍대 문제가 뭔지 이해가 잘 안 되어서 찾아봤습니다.

A = m x n
G = r x n
c = n
b = m
h = r

원문제

$$\begin{aligned} \min_{x} \quad & c^T x \\ subject \quad to \quad & Ax = b \\ & Gx \leq h \end{aligned}$$

u, v 벡터를 양변에 곱한다. (모든 요소 > 0)

$$\begin{aligned} u^T Ax &= u^T b \\ v^T Gx &\leq v^T h \end{aligned}$$

유도

$$\begin{aligned} u^T Ax + v^T Gx &\leq u^T b + v^T h \\ \left(u^T A + v^T G\right) x &\leq u^T b + v^T h \\ \left(A^T u + G^T v\right)^T x &\leq u^T b + v^T h \\ \left(-A^T u - G^T v\right)^T x &\geq -u^T b - v^T h \\ \therefore \quad c^T x &\geq -u^T b - v^T h \end{aligned}$$

쌍대문제

$$\begin{aligned} \max_{u,v} \quad & -b^T u - h^T v \\ subject \quad to \quad & -A^T u - G^T v = c \\ & v \geq 0 \end{aligned}$$

1) https://ratsgo.github.io/convex%20optimization/2018/01/25/duality/

조성진

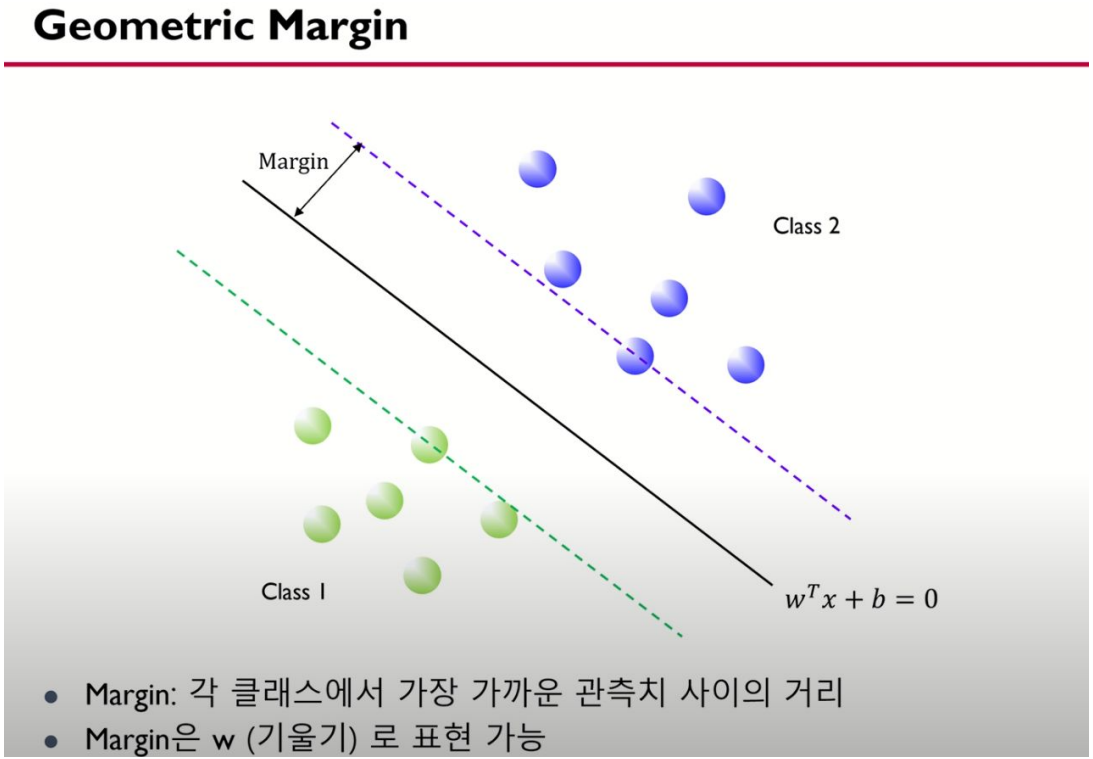
Margin

서포트 벡터

Hard Margin

마진오류

Soft Margin



1) https://www.youtube.com/watch?v=qFg8cDnqYCI&feature=youtu.be

구건모

P-213, 규제에 편향이 포함되면 데이터 정규화는 필수다?

P-209

LinearSVC는 규제에 편향(bias)을 포함시킵니다. 그래서 훈련 세트에서 평균을 빼서 중앙에 맞춰야 합니다. StandardScaler를 사용하여 데이터 스케일을 맞추면 자동으로 이렇게 됩니다.

> (제 직관으로는) 중앙에 맞추면 모델의 편향이 0으로 바뀌기 때문이다.

별개로 Scale을 해야하는 이유가 하나 더 생겼습니다.

1. 각 특성간의 벡터 공간상의 거리를 맞춰 준다.
2. 최적화 알고리즘이 목적값을 찾기위한 경로를 쉽게 만들어준다.
3. Scale을 함으로써 편향에 규제가 적용되는 것을 피할 수 있다.

LinearSVC는 규제에 편향을 포함시키기 때문에 정규화를 진행하지 않을 경우 편향(bias)이 0으로 수렴하게 됩니다. 그러면 실제 데이터 분포와 멀어질 수 있습니다.

`StandardScalar'를 사용하여 정규화를 진행하게 되면 통계값을 저장할 수 있으므로 학습 후 원래 데이터 분포로 쉽게 변환할 수 있습니다.

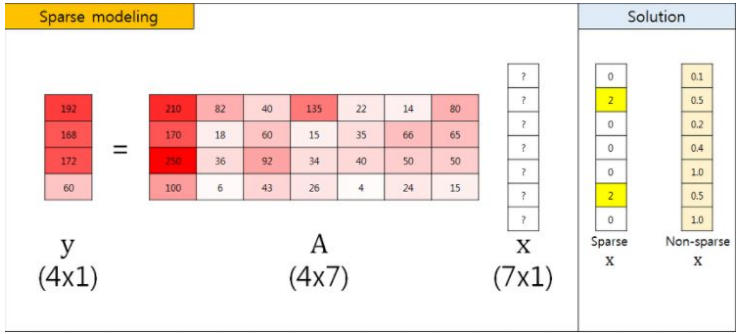
조성진

희소(sparse) 행렬, 밀집(dense) 행렬

행렬의 값이 대부분 0인 경우를 가리키는 표현 밀집 행렬은 반대되는 표현

희소(sparse) 모델

주로 0으로 이루어진 특성의 가중치가 높은 모델.
-> 라쏘 회귀는 희소 모델을 만든다.



1) https://dp-story.tistory.com/4

희소(sparse) 특성

희소 데이터가있는 특성은 대부분 값이 0 인 특성입니다.

희소 특성의 예로는 원-핫 인코딩 된 단어의 벡터 또는 범주 형 데이터 개수가 있습니다.

밀집(dense) 특성
- 일반적으로 부동 소수점 값의 텐서로 이루어진 특성

1) https://www.kdnuggets.com/2021/01/sparse-features-machine-learning-models.html

구건모

LinearSVC 정밀도를 높이면 알고리즘의 수행 시간이 길어지는 이유는 뭘까요?

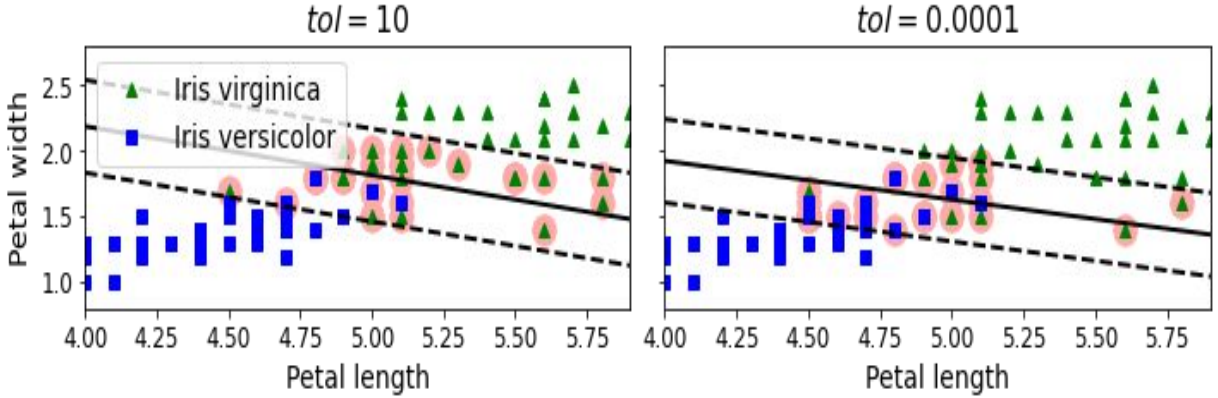
여기서의 정밀도는 분류기의 정밀도와는 다른 개념으로 받아들이는 것이 이해하기 쉽습니다.

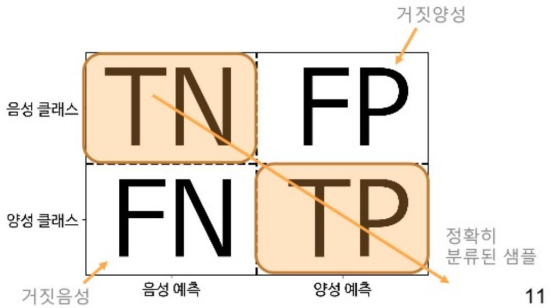
정밀도를 높이는 것은 허용오차를 낮추는 것을 의미하는데 허용오차를 낮추기위해서 더 많은 학습을 진행해야 하므로 알고리즘의 수행 시간이 더 길어집니다.

P-215

정밀도를 높이면 알고리즘의 수행 시간이 길어집니다.

허용오차 하이퍼파라미터 tol 로 조절한다. **(양성 샘플: 세모)**





질문(?)

김유지

||w||는 l2 norm이다?

p. 221

주석 18에서 ||w||는 l2 norm을 말하고, l2 norm의 미분은 원점에서 분모가 0이 되므로 도함수가 정의되지 않는다고 했는데
이는 l1 norm의 설명이 아닌가요?
오타인가 제가 잘못 알고 있는 것인가 궁금합니다

오타였습니다 :)

조성진

p215 희소 특성 아래 두줄

특히 희소 특성(각 샘플에 0이 아닌 특성이 몇 개 없는 경우)인 경우에는 잘 확장됩니다.
이런 경우 알고리즘의 성능이 샘플이 가진 0이 아닌 특성의 평균 수에 거의 비례합니다.

잘 확장된다? 빠르다는 걸까요?

샘플이 가진 0이 아닌 특성의 평균 수 -> 희소 특성의 평균 수??

결론이 잘 나지 않은 질문입니다. 가장 유력한 가설은 아래와 같습니다.

희소 행렬일 경우 벡터의 차원수와 상관 없이 0이 아닌 값들만 계산을 하면 됩니다. 그래서 잘 확장된다는 것을 의미하는 것 같습니다.

위에서 말하는 성능이 알고리즘 속도라면 위와 같은 이유로 설명이 가능합니다. 만약 지표에 의한 성능이라면 이유는 잘 모르겠습니다.

질문

백서윤

p208

[그림 5-4] 의 오른쪽 그래프가 이 코드로 만들어진 것입니다.

c = 1 인데 왜 오른쪽인지, 왼쪽이 아닌가요?

넵, 그림이 잘못 삽입됐습니다.