

# 기계학습 서포트 벡터 머신 정리노트

## (4조 의견)

### [ 하드마진 분류, 소프트마진 분류 ]

하드마진 분류는 데이터가 선형적으로 구분될 수 있어야 제대로 작동하고, 이상치에 민감하기 때문에 그것보다 도로의 폭을 가능한 한 넓게 유지하는 것과 마진 오류 사이에 적절한 균형을 잡는 소프트마진 분류를 사용하는 것이 더 좋아 보인다고 판단된다.

## (4조 의견)

### [ 비선형 SVM ]

비선형 데이터셋을 다룰 때는 다항 특성과 같은 특성을 더 추가해야 한다.

예를들어,  $x_1$ 만 가진 간단한 데이터셋은  $x_2 = (x_1)^2$ 을 추가하여 만들어 2차원으로 만들어 완벽하게 선형적으로 구분할 수 있다.

## (4조 의견)

### [ 다항식 커널 ]

다항식 특성을 추가하는 것은 간단하고 모든 머신러닝 알고리즘에서 잘 작동한

다. 하지만, 낮은 차수의 다항식은 매우 복잡한 데이터셋을 잘 표현하지 못한다. 높은 차수의 다항식은 굉장히 많은 특성을 추가하므로 모델을 느리게 한다.

코드에서 보면 `coef0`의 값을 조정하여 다항식의 차수를 줄일 수 있고, 이를 통해, 과대적합, 과소적합에 대응할 수 있다.

#### (4조 의견)

##### [ 가우시안 RBF 커널 ]

Gamma와 C를 바꾸어서 훈련시키는 모델

Gamma를 증가시키면 종 모양 그래프가 좁아져서 각 샘플의 영향 범위가 작아진다. 이것은 결정경계가 불규칙해지고 각 샘플에 따라 구불구불하게 휘어진다.

반대로, gamma를 감소시키면 종 모양 그래프를 만들며 샘플이 넓은 범위에 걸쳐 영향을 주어 결정 경계가 더 부드러워진다.

하이퍼파라미터  $r$ 가 규제 역할을 한다. 이를 통해 과대적합, 과소적합 대응이 가능하다.