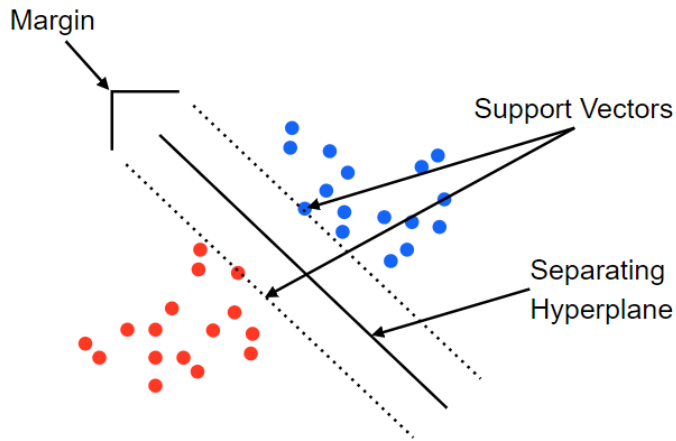


SVM (소프트벡터머신)

*SVM 한줄요약-> 마진을 최대로 하는 지점과 마진오류 사이의 적절한 균형을 잡을 수 있는 지점을 찾아 데이터를 나누는 직선을 긋는 모델



- 선형, 비선형, 분류, 회귀, '이상치탐지'에 사용

- **서포트벡터**: 도로 경계에 위치한 샘플

- 분포가 다른 데이터의 경우 마진의 값이 너무 작아지는 문제가 생기기 때문에 **스케일링**을 하는 것이 좋음

- 속도가 빠른 모델은 아니기 때문에 **작거나 중간크기의 데이터에 적합**

- 하드 마진 VS 소프트 마진

이름	특징
하드마진	<ul style="list-style-type: none"> - 데이터를 나누는 선을 그을 때, 모든 샘플이 도로 바깥쪽에 올바르게 분류되도록 도로를 설정하는 것 - 이상치에 민감함 - 이상치에 민감하기 때문에 일반화가 쉽지 않음 -> 과적합의 문제
소프트마진	<ul style="list-style-type: none"> - 하드마진보다 유연한 모델 - 이상치에 유연함 (일부 데이터는 허용하면서 도로를 설정)

➔ 도로의 폭(마진)을 넓게 유지하는 것과 마진오류 사이의 적절한 균형을 잡는 것이 중요

*마진오류: 샘플이 도로 중간이나 아예 반대편에 있는 경우

비선형 SVM

- 유사도 특성

-> 비선형 특성을 다루는 기법 중 하나로 각 샘플이 특정 랜드마크와 얼마나 닮았는지를 유사도 함수로 계산해 그 특성을 추가하는 방법

-> 0부터 1까지의 값을 가질 수 있으며 종모양임.

-> 랜드마크는 모든 샘플의 위치에 지정 -> 훈련세트가 큰 경우 아주 많은 특성이 생기는 단점

-> **가우시안 RBF커널** 사용