SVM(Support Vector Machine)

SVM은 샘플의 특성으로부터 모델링하고 결정 경계를 찾는 방식이 아닌, 결정경계 자체에 초점을 맞춤.
결정경계는 차원에 따라 형태가 변하지만 여전히 선형적이어서 이를 초평면이라고 칭함.
SVM은 초평면과 Support Vector와의 거리인 margine을 최대화 하고자 함. 정확히는 모든 차원에 있어 최소의 마진이 최대화되는 방향으로 초평면을 찾음.

→ maximize M.

$y_i(\beta_0+\beta_1 x_{i1} \cdots \beta_p x_{ip}) \geq M \ \forall i=1 \cdots n$

→ 스케일을 고정하기 위한 제약조건도 있음

$\sum_{j=1}^{p} \beta_j^2 = 1.$

소프트 마진과 하드 마진

하드 마진의 경우 적절한 결정경계를 찾을 수 없는 경우가 있음

→ 

소프트 마진은 일부의 에러를 용인하는 개념임. 잘못 분류된 샘플에 대해서는 음의 마진을 가짐.
소프트 마진의 최적화 공식은 아래와 같음

→ maximize M

$y_i(\beta_0+\beta_1 x_{i1} \cdots \beta_p x_{ip}) \geq M(1-\epsilon_i),$

$\epsilon_i \geq 0, \ \sum_{i=1}^{n} \epsilon_i \leq C$

→ 즉 C에 의해 허용할 음의 마진 갯수를 조절할 수 있음
C가 커질수록 모델의 유연성은 증가함.

SVM에서 비선형 결정경계

고차항을 도입하여 비선형성을 포착할 수 있으나 실제로 고차항을 만드는 방법보다는 커널을 사용하여 일반화 함.
SVM은 $\hat{L}(\alpha)$를 최소화하는 최적화 모델이나 모든 항을 고차항으로 투영시키는 방법 대신 $\sum_{k=1}^{p} x_{ik}x_{jk}$ 만으로 동일한 효과를 찾을 수 있음

→ Linear kernal

$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_i x_{i'}$

내부 곱셈만으로 커널을 구현 함.

→ Polynomial Kernel

$K(x_i, x_{i'}) = (1+\sum_{j=1}^{p} x_{ij}x_{i'j})^d$

내부 곱셈에 d차항을 도입하는 개념으로서 d에 따라 유연성이 결정됨.

→ Radical kernel (가우시안 커널)

$K(x_i, x_{i'}) = \exp(-\gamma\sum_{j=1}^{p} (x_{ij}-x_{i'j})^2)$

두 벡터에 대해 유클리디안 거리를 구하고, 감마(γ)를 곱해 지수의 형태로 표현함.
가장 많이 사용되며, 감마에 의해서 유연성이 결정됨.