(1). Support vector machine(SVM)

SVM은 서로 다른 집단에 속한 범주형 데이터를 분류하는 최적의 초평면을 찾는 모형이다. 먼저 SVM을 훈련하는데 이용하는 데이터 집합 D를 다음과 같이 정의한다.

$$D = \left\{(x_i, y_i) | x_i \in \mathbb{R}^m, y_i \in \{-1, 1\}\right\}_{i=1}^{n}$$

각 x_i는 m차원의 실수 벡터를 나타내며 y_i는 x_i가 어떤 범주에 속해있는지에 따라 1 또는 -1의 값을 가진다. 다음으로 y_i의 값에 따라 데이터를 기하학적으로 분리할 수 있을 때, 그 경계면을 초평면(hyperplane)이라 부르며 초평면과 가장 가까운 데이터들을 서포트 벡터(support vector)라고 정의한다. SVM은 이러한 서포트 벡터들을 바탕으로 두 분류 집단 사이의 공백(margin)을 최대화하는 초평면을 찾으며, 그 구조를 Figure 1과 같이 나타낼 수 있다.

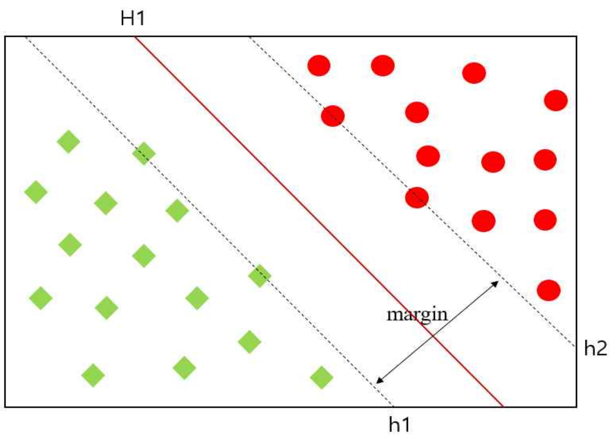

Figure 1. Structure of SVM.

두 분류 집단을 구분하는 선형 경계면을 $f(x) = w^T x + b$으로 정의하면 최적의 경계면을 찾는 문제는 두 분류 집단의 서포트 벡터들 사이의 거리인 $\frac{2}{\| w \|^2}$을 최대화하는 문제로 나타낼 수 있는데, 이는 $\frac{1}{2}\| w \|^2$을 최소화하는 것과 동치다. 마지막으로 각 집단에 속하는 데이터들이 경계면을 기준으로 같은 방향에 위치하도록 $y_i(w^T x + b) \geq 1$의 제약조건을 적용하여 비용함수를 최소화하는 w와 b를 계산하면 최적의 경계면을 찾을 수 있으며, 이를 수식으로 표현하면 다음과 같다.

$$\min_{w,b} \frac{1}{2} \parallel w \parallel^2$$
$$s.t.\ y_i(w^T x_i + b) \geq 1 \quad \text{for } i = 1, \cdots, N$$

위의 SVM 모형은 선형 경계면에 의해 데이터가 완전하게 두 집단으로 분류되는 경우를 산정하고 있다. 하지만 이 모형은 데이터가 완전하게 선형 분리가 되지 않는 데이터들을 다루는 경우에는 적용할 수 없는데, 이 경우에는 유화변수(slack variable) ζ_i과 패널티 변수(penalty variable) C를 도입하여 해결한다.

$$\min_{w,b} \frac{1}{2} \parallel w \parallel^2 + C\sum_{i=1}^{N}\zeta_i$$
$$s.t.\ y_i(w^T x_i + b) \geq 1 - \zeta_i\ ,$$
$$\zeta_i \geq 1 \quad \text{for } i = 1, \cdots, N$$

ζ_i는 데이터가 SVM의 마진 안에 위치할 수 있도록 해주는 역할을 하는 반면, C는 마진 안에 위치한 데이터들에 대해 부과하는 패널티로 이해할 수 있다. 이를 그림으로 나타내면 Figure 2와 같다.

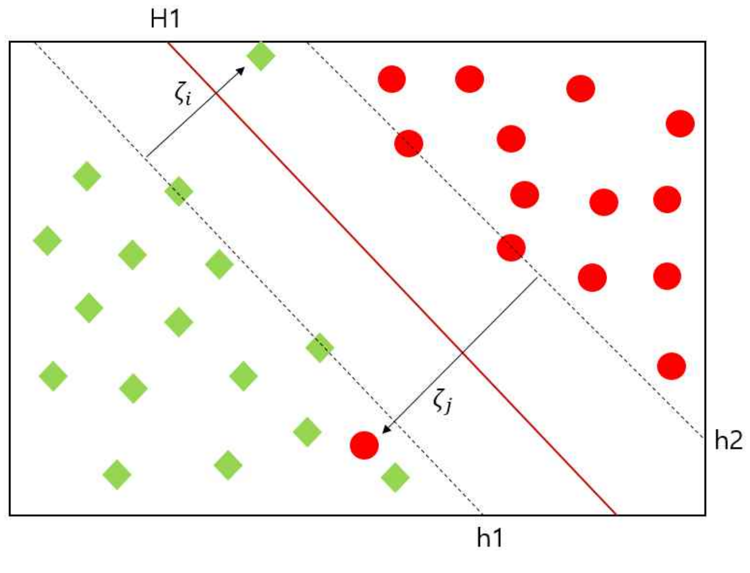

Figure 2. Structure of SVM with ζ and C.

마지막으로 두 분류 집단을 나누는 초평면 자체가 비선형인 경우가 있다. 이 경우에는 커널 함수(Kernel function) $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$을 이용해 데이터를 원 공간(input space)으로부터 선형으로 분류할 수 있는 고차원의 특징 공간(feature space)으로 투사한 뒤 두 집단을 나누는 초평면을 찾는다. 이러한 커널-SVM(Kernel-SVM)을 그림으로 나타내면 Figure 3과 같다. 커널-SVM에 적용하는 커널 함수는 다양한 형태를 가질 수 있는데, 기존 연구들에서 가장 많이 사용된 커널 함수들은 다음과

같다.

$$
\begin{aligned}
Linear &: K(x_i,x_j) = x_i^T x_j, \\
Polynomial &: K(x_i,x_j) = (x_i^T x_j + c)^d,\ c > 0, \\
Sigmoid &: K(x_i,x_j) = \tanh\{a(x_i^T x_j) + b\},\ a,b \geq 0, \\
Gaussian &: K(x_i,x_j) = \exp - \frac{\| x_i - x_j \|^2}{2\sigma^2}, \sigma \neq 0.
\end{aligned}
$$

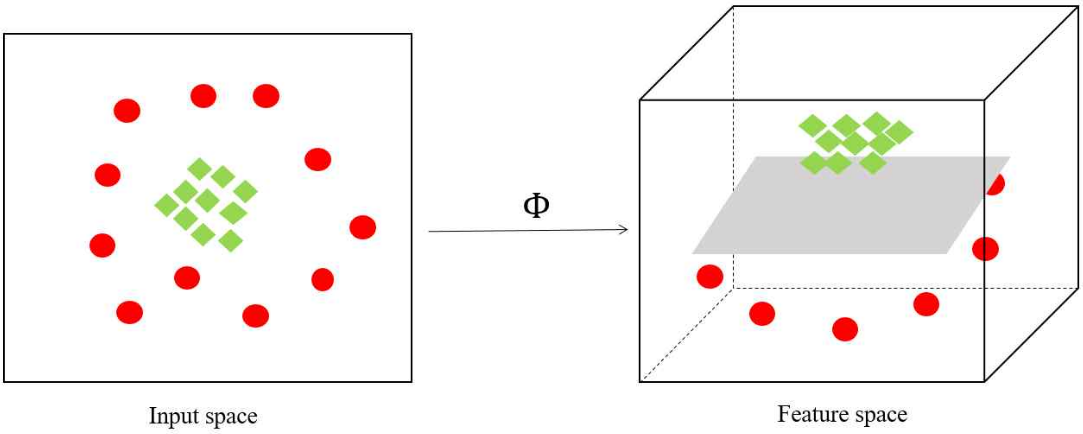

Figure 3. Structure of Kernel-SVM.