SVM-KKT

SVM을 최적화 하기 위해선 KKT 조건을 활용한다.

SVM의 제약조건은 릿지회귀에서 보다 더 복잡하므로 KKT 조건을 사용하여 최적화를 진행한다.

목적함수 $\min \frac{1}{2}\|w\|^2$

제약조건 $y_i(w^T x_i + b) \ge 1$

라그랑주 함수 정의

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \{ y_i(w^T x_i + b) - 1 \}$$

KKT 조건

① stationarity

$$\nabla L(w, b, \alpha) = 0$$

② primal feasibility

$$y_i(w^T x_i + b) \ge 1$$

③ dual feasibility

$$\alpha_i \ge 0$$

④ complementary slackness

$$\alpha_i \{ y_i(w^T x_i + b) - 1 \} = 0$$

Stationarity

$\nabla L(w,b,\alpha)$를 계산해보겠다.

$$L(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \{ y_i (w^T x_i + b) - 1 \}$$

$$\frac{d}{dw} L(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$\frac{d}{db} L(w,b,\alpha) = - \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

종합하자면, 다음 2가지 결과를 얻었다

$$w^* = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

라그랑주 함수를 duality를 이용하여 푸는 이유는 다음과 같다.

라그랑주 함수의 개형을 보면 아래와 같은데

$$L(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i y_i (w^T x_i + b) - 1$$

w와 b에 대해서 라그랑주 함수는 convex하다. 함수가 convex하다는 것은 최적화 문제에서 큰 이점을 갖는데, 그 이유는 Convex 함수의 경우, local minimum이 곧 global minimum이 되며, convex 함수의 stationary point(그래디언트가 0이 되는 지점)에서 최적화를 제공하는 강력한 이점이 있다.

하지만, 또 다른 변수인 승수 α에 대해선 convex하지 않다. 즉, 목적함수를 최소화 하려면 w와 b를 최소화 해야 하지만 α는 최대화 하는 문제가 있다.

따라서 duality를 이용하여 라그랑주 함수를 α에 대한 식으로만 나타내어서 α를 최대화 하여 최적화를 이루어 낸다.

dual problem의 변환은 다음과 같은 이점을 제공한다.

1. 계산 효율성
   듀얼 문제는 종종 원 문제보다 더 적은 수의 변수로 표현될 수 있다. 특히, SVM에서는 훈련 데이터의 수가 특성의 수보다 많을 때 이 이점이 두드러진다.
2. kernel trick 사용
   듀얼 문제 형태를 사용하면, kernel trick을 적용하여 고차원 공간에서의 선형 분리를 용이하게 할 수 있다. 이는 비선형 분류 문제를 해결하는 데 효과적이다.
3. sparse 해
   듀얼 문제의 해는 대부분의
   α가 0이 되는 sparse해(희소 해)를 제공한다. 이는 결정 경계를 정의하는 데 단지 소수의 서포트 벡터만이 필요하다는 것을 의미한다. 이로 인해 모델의 일반화 능력이 향상되고, 계산 비용도 절감된다.

$L(w, b, \alpha)$를 $w^* = \sum_{i=1}^{n} \alpha_i y_i x_i$ 를 이용하여

dual 문제 $L(w^*, b^*, \alpha)$로 바꾸어 풀어보겠다.

$$\frac{1}{2}\|w^*\|^2 = \frac{1}{2} w^{*T} w^*$$

$$= \frac{1}{2} w^{*T} \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i \left(w^{*T} x_i \right)$$

$$w^{*T} = \sum_{j=1}^{n} \alpha_j y_j x_j^T$$

$$= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i \sum_{j=1}^{n} \alpha_j y_j x_j^T x_i$$

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i$$

$$\sum_{i=1}^{n} \alpha_i \left\{ y_i (w^{*T} x_i + b) - 1 \right\}$$

$$= \sum_{i=1}^{n} \alpha_i y_i (w^{*T} x_i + b) - \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i y_i w^{*T} x_i + b \sum_{i=1}^{n} \alpha_i y_i - \sum_{i=1}^{n} \alpha_i$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\sum_{i=1}^{n} \alpha_i y_i w^{*T} x_i = \|w^*\|^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i - \sum_{i=1}^{n} \alpha_i$$

이를 통해서 $L(w^*, b^*, \alpha) = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i$

원래 라그랑주 함수는 $L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \{ y_i (w^T x_i + b) - 1 \}$

목적함수인 $\frac{1}{2}||w||^2$을 최소화하려면 α를 최대화 해야 한다.

$L(w^*, b^*, \alpha)$를 $L_{dual}(\alpha)$로 둔다면 이를 최대화 해야 한다.

정리하자면,

$$\text{maximize } L_{dual}(\alpha) = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_j^T x_i$$

$$\text{subject to } \begin{cases} \sum_{i=1}^{n} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases}$$