Constrained Optimization

1. Equality Constraints : 등식 제한 조건
   1) $f(x)$를 최소화하고
   2) $g(x)=0$를 만족

2. Inequality Constraints : 부등식 제한 조건
   1) $f(x)$를 최소화하고
   2) $g(x) \le 0$를 만족

   여기서 $f(x)$를 Cost Function이라고 생각

제약조건이 Active
$\rightarrow g(x)=0$, 경계에 존재
제약 조건이 Inactive
$\rightarrow g(x)<0$, 경계 안쪽에 존재
$\Rightarrow$ 제약 조건이 있으나 없으나 의미X
$\lambda=0$으로 설정해서 제약 조건을 무시

3. 라그랑주 승수법
   등식 제한 조건 or Active한 경우

$f(x_1, x_2) = x_1^2 + x_2^2$
$g(x_1, x_2) = x_1 + x_2$
$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2)$

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0$$
$$\frac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 = 0$$
$\Rightarrow (x_1, x_2) = (0.5, 0.5)$

Inactive 한 경우
제약 조건이 없다고 생각하고 그냥 최적화

4. KKT 조건

$\nabla_{\vec{x}} L(\vec{x}, \mu) = 0$
$g(x) \le 0$
$\lambda \ge 0$
$\lambda g(x) = 0$

Active 한 경우
$g(x)=0, \lambda>0$ / $\nabla_\lambda L = 0$
Inactive 한 경우
$g(x)<0, \lambda=0$

따라서 문제를 풀 때에는 $g(x)=0$ (Active), $\lambda=0$ (Inactive)인 모든 경우를 고려하여 풀이, Inactive한 경우 $g(x)<0$을 만족하고 결과는 Inactive한 경우를 따라감

5. Multiple Constraints : 제약 조건이 여러 개
   그냥 제약 조건 개수만큼 승수가 있음

$f(x)$
$g(x)=0$
$h(x)=0$
$L(x, \lambda, \nu)$

6. Largrangian Duality
   원래문제(primal)를 새로운 문제(dual)로 변환

primal: $\min_{\vec{x}} f(\vec{x}), \; g_i(\vec{x}) \leqslant \vec{0} \; (i=1,2,\cdots,m)$

dual: $\max_{\substack{\vec{\lambda} \in \mathbb{R}^m \\ \vec{\lambda} \geqslant \vec{0}}} D(\vec{\lambda}) = \min_{\vec{x} \in \mathbb{R}^d} \mathcal{L}(\vec{x}, \vec{\lambda}) \longrightarrow$ 라그랑주 함수를 최소로 만드는 $\vec{x}$

$\nabla_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) = 0 \Rightarrow \vec{x}^*$
$D(\vec{\lambda}) = \mathcal{L}(\vec{x}^*, \vec{\lambda})$

Ex) Primal: $\min_{\vec{x}\in\mathbb{R}^d} \vec{c}^T\vec{x}$ (Linear)

$\vec{A}\vec{x}\le\vec{b}$

$\Rightarrow \mathcal{L}(\vec{x},\vec{\lambda}) = \vec{c}^T\vec{x}+\vec{\lambda}^T(\vec{A}\vec{x}-\vec{b})$

$=(\vec{c}+\vec{A}^T\vec{\lambda})^T\vec{x}-\vec{\lambda}^T\vec{b}$

$\nabla_{\vec{x}}\mathcal{L}(\vec{x},\vec{\lambda})=\vec{c}+\vec{A}^T\vec{\lambda}=\vec{0}$

$D(\vec{\lambda})=-\vec{\lambda}^T\vec{b}$

Dual: $\max_{\vec{\lambda}\in\mathbb{R}^m} -\vec{b}^T\vec{\lambda}$

$\vec{c}+\vec{A}^T\vec{\lambda}=\vec{0}$

$\vec{\lambda}\ge\vec{0}$

Primal: $\min_{\vec{x}\in\mathbb{R}^d} \frac{1}{2}\vec{x}^T\vec{Q}\vec{x}+\vec{c}^T\vec{x}$ (Quadratic)

$\vec{A}\vec{x}\le\vec{b}$

$\mathcal{L}(\vec{x},\vec{\lambda})=\frac{1}{2}\vec{x}^T\vec{Q}\vec{x}+\vec{c}^T\vec{x}+\vec{\lambda}^T(\vec{A}\vec{x}-\vec{b})$

$=\frac{1}{2}\vec{x}^T\vec{Q}\vec{x}+(\vec{c}+\vec{A}^T\vec{\lambda})^T\vec{x}-\vec{\lambda}^T\vec{b}$

$\nabla_{\vec{x}}\mathcal{L}(\vec{x},\vec{\lambda})=\vec{Q}\vec{x}+\vec{c}+\vec{A}^T\vec{\lambda}=0$

$\vec{x}=-\vec{Q}^{-1}(\vec{c}+\vec{A}^T\vec{\lambda})$

Dual: $\max_{\vec{\lambda}\in\mathbb{R}^m} -\frac{1}{2}(\vec{c}+\vec{A}^T\vec{\lambda})\vec{Q}^{-1}(\vec{c}+\vec{A}^T\vec{\lambda})-\vec{\lambda}^T\vec{b}$

$\vec{\lambda}\ge\vec{0}$

Support Vector Machine

1. Margin (여백)

결정 초평면에서 양쪽 class에 대한 여백을 최대화할수록 일반화 능력이 향상될 것이다

2. Linear SVM

Linear SVM에서의 Margin

$\text{Margin}=2s=\frac{2|d(x)|}{\|w\|_2}=\frac{2}{\|w\|_2}$

d(x) (결정 초평면)에 상수를 곱해도 변하지 않는 성질을 이용해 $|d(x)|=1$

Maximize: $J(w)=\frac{2}{\|w\|_2}$

Subject to: $\vec{w}^T\vec{x}_i+b\ge0, \forall y_i=1$

$\vec{w}^T\vec{x}_i+b\le0, \forall y_i=-1$

↓ 역수

Minimize: $J(w)=\frac{\|w\|_2}{2}$

Subject to: $y_i(\vec{w}^T\vec{x}+b)-1\ge0$

$\mathcal{L}(\vec{w},b,\vec{\alpha})=\frac{\|w\|_2}{2}-\sum_i^n\alpha_i(y_i(\vec{w}^T\vec{x}+b)-1)$

KKT

$\frac{\partial\mathcal{L}(\vec{w},b,\vec{\alpha})}{\partial\vec{w}}=\vec{0},\ \vec{w}=\sum_i^n\alpha_i y_i\vec{x}_i$

$\frac{\partial\mathcal{L}(\vec{w},b,\vec{\alpha})}{\partial b}=0,\ \sum_i^n\alpha_i y_i=0$

$\alpha_i\ge0$

→ 결국 $\vec{\alpha}_i$와 Support Vector를 구하면 됨

$\vec{\alpha}_i$ 와 $\vec{x}_i$를 구하면 $\vec{w}$ 구할 수 있고,

$\vec{w}$ 와 $\vec{x}_i$ 를 구하면 b 구할 수 있다.

3. Support Vector

$y_i(\vec{w}^T\vec{x}_i+b)-1=0$ 을 만족하는 $\vec{x}_i$

4. Wolfe Dual Problem

Maximize : $\tilde{L}(\alpha) = \sum_{i}^{n} \alpha_i - \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n} \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$

Subject to : $\sum_{i}^{n} \alpha_i y_i = 0$

$\alpha_i \geq 0, \ i = 1, 2, \cdots, n$

5. Soft Margin

Hard Margin과 달리 오분류 샘플을 허용

Slack 변수 도입

$y_i(\vec{w}^T \vec{x}_i + b) \geq 1 - \xi_i$

Minimize : $J(\vec{x}, \xi) = \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i}^{n} \xi_i$

C는 하이퍼 파라미터

C↓ 여백의 비중↑ 여백 최대화

C↑ ξ의 비중↑ 오분류 샘플 최소화

Subject to : $y_i(\vec{w}^T \vec{x}_i + b) \geq 1 - \xi_i$

$0 \leq \xi_i, \ i = 1, 2, \cdots, n$

Largrangian

$\mathcal{L}(\vec{w}, b, \xi, \vec{\alpha}, \vec{\beta}) = \left(\frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i}^{n}\xi_i\right) - \left(\sum_{i}^{n} \alpha_i (y_i(\vec{w}^T \vec{x}_i + b) - 1 + \xi_i) + \sum_{i}^{n} \beta_i \xi_i\right)$

KKT

$\frac{\partial L}{\partial \vec{w}} = \vec{0}, \ \vec{w} = \sum_{i}^{n} \alpha_i y_i \vec{x}_i$

$\frac{\partial L}{\partial b} = \vec{0}, \ \sum_{i}^{n} \alpha_i y_i = 0$

$\frac{\partial L}{\partial \xi_i} = \vec{0}, \ C = \alpha_i + \beta_i, \ i = 1, 2, \cdots, n$

6. Radial Basis Function

$\phi(\vec{x}) = \|\vec{x} - \vec{c}\|$

⇒ 어떤 고정점 c에서 입력 벡터 x까지의 거리에만 영향을 받는 함수

ex) Gaussian Function

$\phi(\vec{x}) = \exp\left(-\frac{\|\vec{x} - \vec{c}\|_2^2}{2\sigma^2}\right)$

7. Traditional Method

공간변환 + 선형 SVM을 통해 학습

↓ Nowadays

Kernal Function 이용

8. Kernal Function

$K(\vec{x}, \vec{z}) = \phi(\vec{x}) \cdot \phi(\vec{z})$ 인 ϕ 존재하면 K는 커널 함수

$\phi : L \to H$ (공간 맵핑)

$K : L \times L \to R$ (커널 함수)

자주 쓰이는 커널 함수

Polynomial : $K(\vec{x}, \vec{z}) = (\vec{x} \cdot \vec{z} + 1)^p$

Gaussian

Hyperbolic Tangent : $K(\vec{x}, \vec{z}) = \tanh(\alpha \vec{x} \cdot \vec{z} + \beta)$

K_1, K_2가 커널 함수라면

$\alpha \cdot K_1 (\alpha > 0), K_1 + K_2, K_1 \cdot K_2$