Ridge Regression-Lagrange Multiplier Method

Using Lagrange multipliers, solve the optimized w

Probabilistic Machine Learning 8.5 내용을 참고하였다.

라그랑주 승수법이란

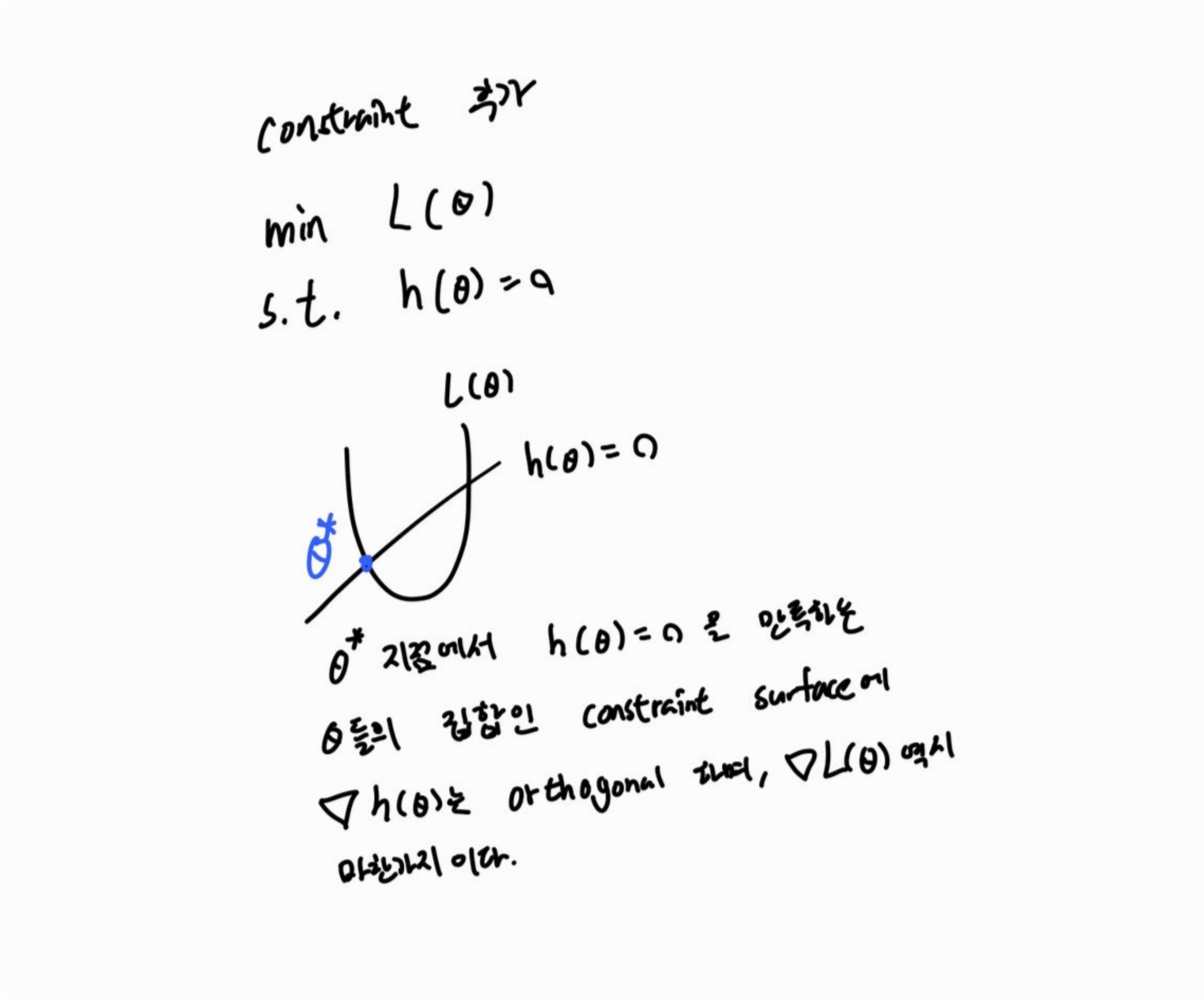
라그랑주 승수법은, 최적화 문제를 푸는 방법중 하나임으로 제약이 있는 최적화 문제에 많이 다뤄진다. 모든 제약식에 라그랑주 승수 λ를 곱하고 등식이없는 문제로 바꾸어 해결하는 방식

$$\nabla L(\theta^*) = \lambda^* \nabla(\theta^*)$$

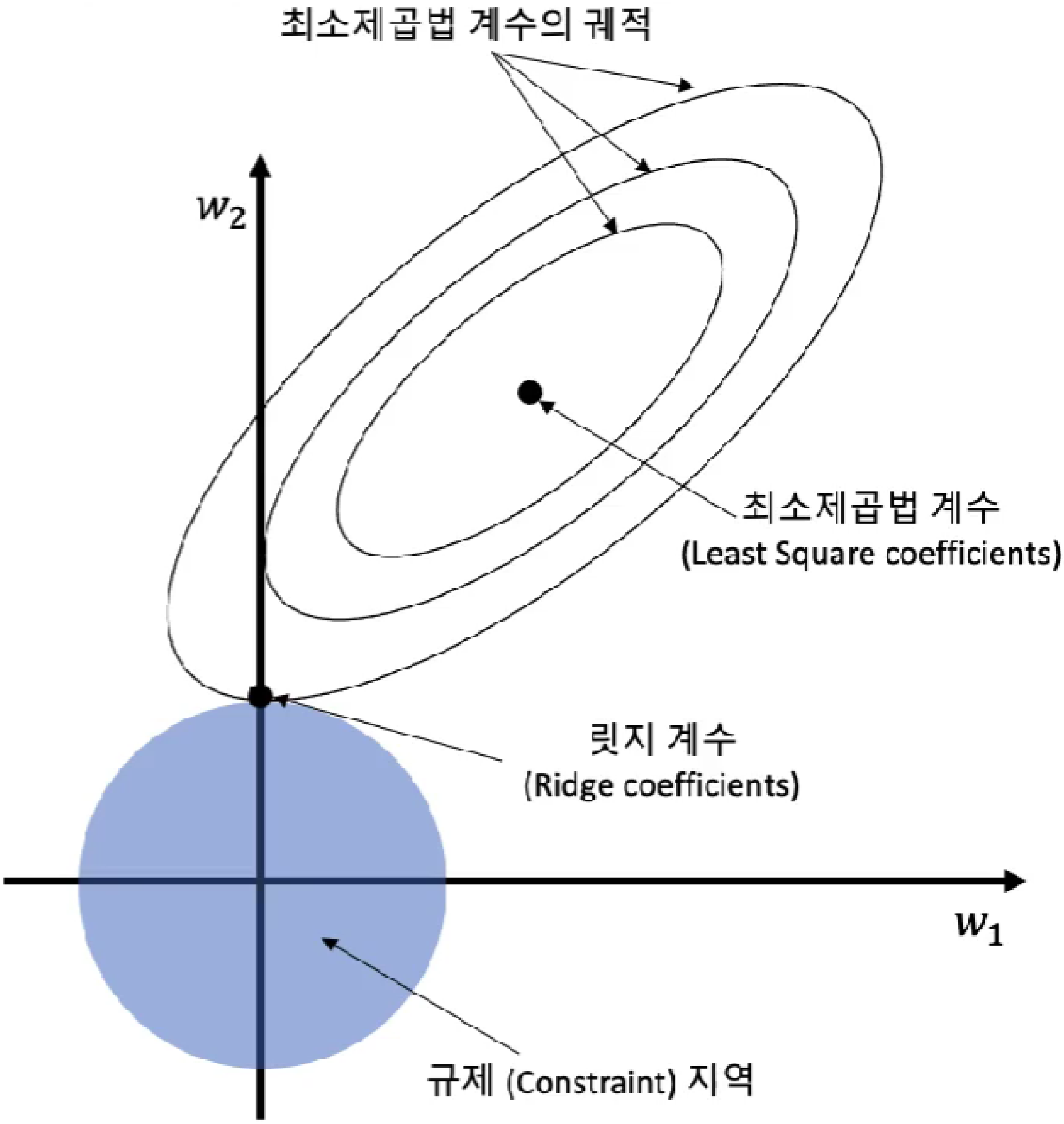
먼저 이 식에 관한 해석이 필요하다.

$\nabla h(\theta)$는 제약조건 함수 $h(\theta)$의 기울기이며, $\nabla L(\theta)$은 목적함수 $L(\theta)$의 기울기를 나타낸다. θ^*는 이러한 제약조건을 만족하는 최적의 파라미터 값이다.

$\nabla h(\theta)$와 $\nabla L(\theta)$는 θ^*에서 constraint surface에 직교한다. 이들은 서로 parallel 상태이거나 anti parallel 상태이다. 즉, 서로 orthogonal하다.



여기서 λ^*는 라그랑주 승수에 해당한다. 이 라그랑주 승수는 양수 혹은 음수가 가능하다.



w_1, w_2 2차원 평면에서 생각해본다면 다음과 같은 의문점이 들었다.

접점인 θ^*에서의 제약조건 함수와 목적함수의 접선의 기울기가 같은데 어떻게 그래디언트가 평행하다는 거지?

목적함수와 제약조건 함수의 접선의 기울기는 같다. 하지만, 이것이 그래디언트가 같다는 의미는 아니다. 이를 좀 더 자세히 이해하기 위해서 접선의 기울기와 그래디언트의 차이를 알아보자.

그래디언트는 고차원 공간에서의 함수의 최대 증가율을 나타내는 벡터이다. 즉, 방향과 크기를 가졌으며, 함수가 가장 크게 증가하는 방향과 그 증가율을 나타낸다.

반면 접선의 기울기는 방향을 가지지 않으며, 그 점에서의 즉각적인 변화율을 나타낸다.

At a stationary point of the Lagrangian, we have

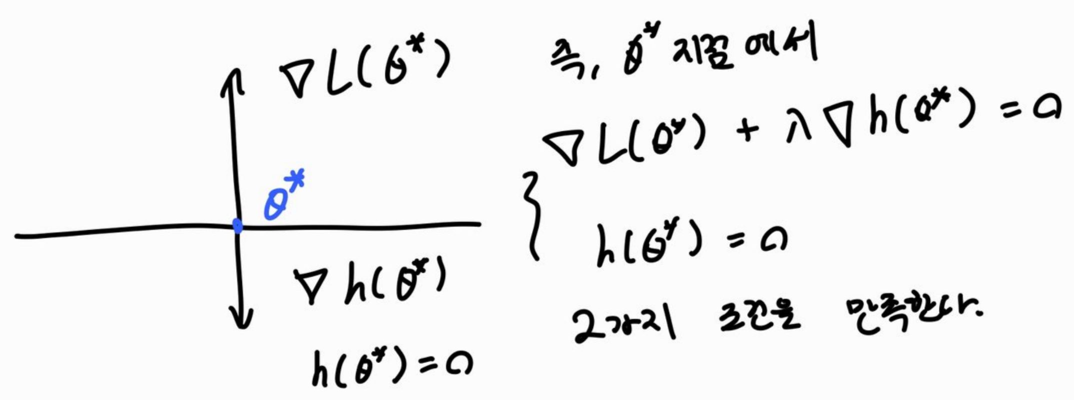
$$\nabla_{\boldsymbol{\theta},\lambda} L(\boldsymbol{\theta}, \lambda) = \mathbf{0} \iff \lambda \nabla_{\boldsymbol{\theta}} h(\boldsymbol{\theta}) = \nabla \mathcal{L}(\boldsymbol{\theta}),\ h(\boldsymbol{\theta}) = 0 \tag{8.91}$$

This is called a **critical point**, and satisfies the original constraint $h(\boldsymbol{\theta}) = 0$ and Equation (8.89).
If we have $m > 1$ constraints, we can form a new constraint function by addition, as follows:

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathcal{L}(\boldsymbol{\theta}) + \sum_{j=1}^{m} \lambda_j h_j(\boldsymbol{\theta}) \tag{8.92}$$

$$\nabla L(\theta, \lambda) = L(\theta) + \sum_{j=1}^{m} \lambda_j h_j(\theta)$$

이제 $\nabla L(\theta, \lambda)$를 정의할 수 있다. 그리고 이 값이 0이 되게 끔 하는 것이 θ를 최적화 하는 것이다.



$\nabla L(\theta, \lambda)$가 0이 된다는 의미는 목적함수와 제약조건 함수의 그래디언트가 반대 방향이 된다.

즉, 서로의 그래디언트가 평행하면서 반대방향이어야 제약조건을 유지하면서 목적함수를 minima 할 수 있다.

아래는 목적함수를 Least_Square_Estimation를 사용하고 목적함수는 L2 노름을 통해서 표현했다.

1. 최소화 해야하는 문제

$$\min (y - Xw)^T (y - Xw)$$

2. 제약 조건

$$\text{subject to } \|w\|_2^2 \le t$$

3. 라그랑주 함수 정의

$$L(w, \lambda) = (y - Xw)^T (y - Xw) + \lambda(\|w\|_2^2 - t)$$

4. 라그랑주 함수를 w나 λ에 대해서 편미분

w에 대한 편미분 부터 진행

$$\frac{dL}{dw} = \frac{d}{dw}\left(y^T y - 2 w^T X^T y + w^T X^T X w + \lambda(\|w\|_2^2 - t)\right)$$

$$= -2X^T y + X^T X w + w^T X^T X + 2\lambda w$$

여기서 $w^T X^T X$는 w와 $X^T X$를 내적한 것이다.

따라서 $w^T X^T X = (w^T X^T X)^T = X^T X w$ 이다.

$$= -2X^T y + 2X^T X w + 2\lambda w$$

$$= 2w(X^T X + \lambda I) - 2X^T y = 0$$

$$\Rightarrow \quad w(X^T X + \lambda I) = X^T y$$

$$w = (X^T X + \lambda I)^{-1} X^T y$$

5. λ에 대한 편미분

$$\frac{dL}{d\lambda} = \frac{d}{d\lambda}\left(\lambda(\|w\|_2^2 - t)\right) = \|w\|_2^2 - t = 0$$

$$\therefore \quad w = (X^T X + \lambda I)^{-1} X^T y$$

$$\|w\|_2^2 = t$$

라그랑주 승주법 적용

1. 최소화 해야하는 문제 (기본 선형 회귀의 목적함수)

$$min(y - Xw)^T(y - Xw)$$

2. 제약 조건

$$subject\ to\ ||w||_2^2 \leq t$$

3. 위 조건으로 라그랑주 함수 정의

$$L(w, \lambda) = (y - Xw)^T(y - Xw) + \lambda(||\ w\ ||_2^2\ -t)$$

4. 라그랑주 함수를 w와 λ에 대해서 편미분 진행
먼저,
w에 대해서 미분해보겠다.

$$\frac{\partial L}{\partial w} = -2X^T(y - Xw) + 2\lambda w = 0$$
$$-2X^Ty + 2X^TXw + 2\lambda w = 0$$
$$= (X^TX + \lambda I)w = X^Ty$$

따라서 최적의 w는 다음과 같이 구할수있다.

$$w = (X^TX + \lambda I)^{-1}X^Ty$$

5. λ에 대해서 미분

$$\frac{\partial L}{\partial \lambda} = ||\ w\ ||_2^2\ -t$$
$$||\ w\ ||_2^2 = t$$

제약조건과 같음으로 별도의 해결과정 없이 제약조건을 그대로 사용한다.

따라서

$$w = (X^TX + \lambda I)^{-1}X^Ty$$
$$||\ w\ ||_2^2 = t$$