

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 λ 를 곱하고 등식이없는 문제로 바꾸어 해결하는 방식

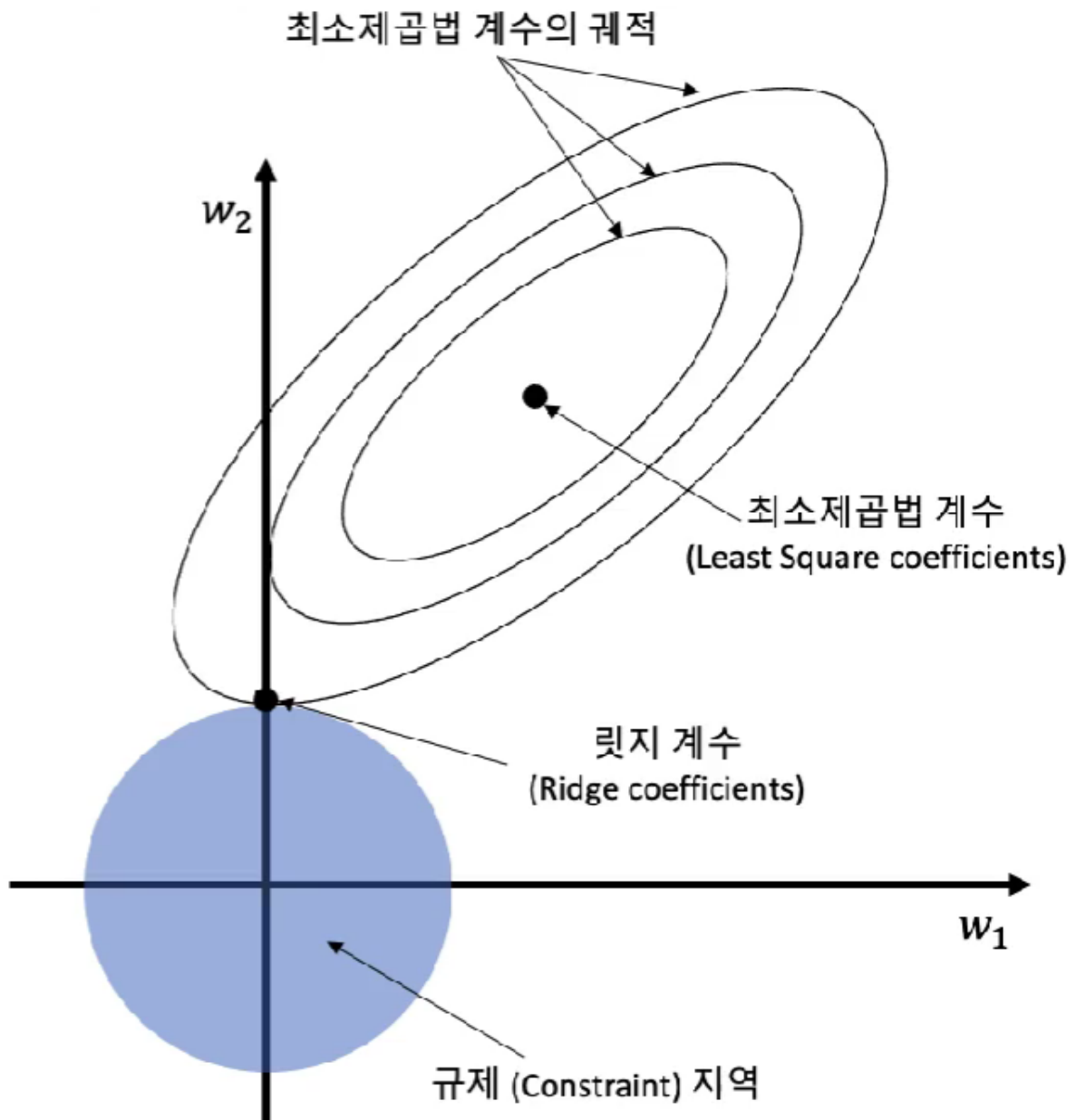
$$\nabla \mathcal{L}(\theta^*) = \lambda^* \nabla h(\theta^*) \quad (8.89)$$

$\nabla h(\theta)$ 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^* 에 직교한다. 여기서 λ^* 는 라그랑주 승수에 해당한다.

직교의 의미

목적함수와 제약조건 함수의 그래디언트가 서로 반대 방향을 가리키고 있으며, 이는 두 벡터가 서로 직교한다는 것을 의미한다.



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

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

목적함수와 제약조건 함수의 접선의 기울기는 같다. 하지만, 이것이 그래디언트가 같다는 의미는 아니다. 이를 좀 더 자세히 이해하기 위해서 "접선의 기울기"를 다시 생각해보자.

"접선의 기울기"라는 말은 고차원에서의 접평면(tangent plane)이나 접선 벡터(tangent vector)로 이해될 수 있으며, 이는 그래디언트 방향을 의미한다.

따라서 접선의 기울기 같으므로 그래디언트의 방향이 같다. 따라서 λ 를 통해서 그래디언트의 방향성을 조절한다.

At a stationary point of the Lagrangian, we have

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

This is called a **critical point**, and satisfies the original constraint $h(\theta) = 0$ and Equation (8.89).

If we have $m > 1$ constraints, we can form a new constraint function by addition, as follows:

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

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

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

즉, 목적함수인 오류함수는 감소하는 방향으로 가고, 제약조건 함수는 증가하는 방향으로 간다.

제약조건 함수가 증가한다는 것은

$\|w\|_2^2$ 가 임계값 (t)보다 크게 되어서 규제항의 영향력이 커진다는 것을 의미한다.

이 경우, 규제항은 가중치 (w)의 값들을 작게 유지하려는 효과를 강화한다. 즉, 모델의 복잡도가 감소하게 된다.

아래는 목적함수를 Least_Square_Estimation를 사용하고 목적함수는 L2 노름을 통해서 표현했다.

1. 최소화 해야하는 문제

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

2. 제약 조건

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

3. 라그랑주 함수 정의

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

4. 라그랑주 함수를 w 과 λ 에 대해서 편미분

w 에 대한 편미분 부터 진행

$$\begin{aligned} \frac{dL}{dw} &= \frac{d}{dw} (y^T y - 2w^T X^T y + w^T X^T X w + \lambda (\|w\|_2^2 - t)) \\ &= -2X^T y + X^T X w + 2\lambda w \end{aligned}$$

여기서 $w^T X^T X$ 는 w 과 $X^T X$ 를 내적한 것이다.

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

$$\begin{aligned} &= -2X^T y + 2X^T X w + 2\lambda w \\ &= 2w(X^T X + \lambda I) - 2X^T y = 0 \end{aligned}$$

$$\begin{aligned} \Rightarrow w(X^T X + \lambda I) &= X^T y \\ w &= (X^T X + \lambda I)^{-1} X^T y \end{aligned}$$

5. λ 에 대한 편미분

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

$$\begin{aligned} \therefore w &= (X^T X + \lambda I)^{-1} X^T y \\ \|w\|_2^2 &= t \end{aligned}$$

라그랑주 승주법 적용

1. 최소화 해야하는 문제 (기본 선형 회귀의 목적함수)

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

2. 제약 조건

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

3. 위 조건으로 라그랑주 함수 정의

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

4. 라그랑주 함수를 w 와 λ 에 대해서 편미분 진행

먼저,
 w 에 대해서 미분해보겠다.

$$\begin{aligned}\frac{\partial L}{\partial w} &= -2X^T(y - Xw) + 2\lambda w = 0 \\ -2X^T y + 2X^T Xw + 2\lambda w &= 0 \\ &= (X^T X + \lambda I)w = X^T y\end{aligned}$$

따라서 최적의 w 는 다음과 같이 구할수있다.

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

5. λ 에 대해서 미분

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

제약조건과 같음으로 별도의 해결과정 없이 제약조건을 그대로 사용한다.

따라서

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