4. 모델 훈련

정윤서

4-1. 선형 회귀

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

- $\hat{y}$: 예측값
- n : 특성의 수
- x_i : i번째 특성값
- θ_j : j번째 모델 파라미터(편향 θ_0 와 특성의 가중치 $\theta_1, \theta_2, \ldots,$ 포함

4-1. 선형 회귀

$$\hat{y} = h_\theta(x) = \theta \cdot x$$

- θ : 편향 θ_0와 θ_1에서 θ_n까지의 특성 가중치를 담은 모델의 파라미터 벡터
- x : x_0에서 x_n까지 담은 샘플의 특성 벡터. x_0는 항상 1이다.
- $\theta \cdot x$: 벡터 두 개의 점곱이다. 이는 $\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$와 같다.
- h_θ : 모델 파라미터 θ를 사용한 가설 함수

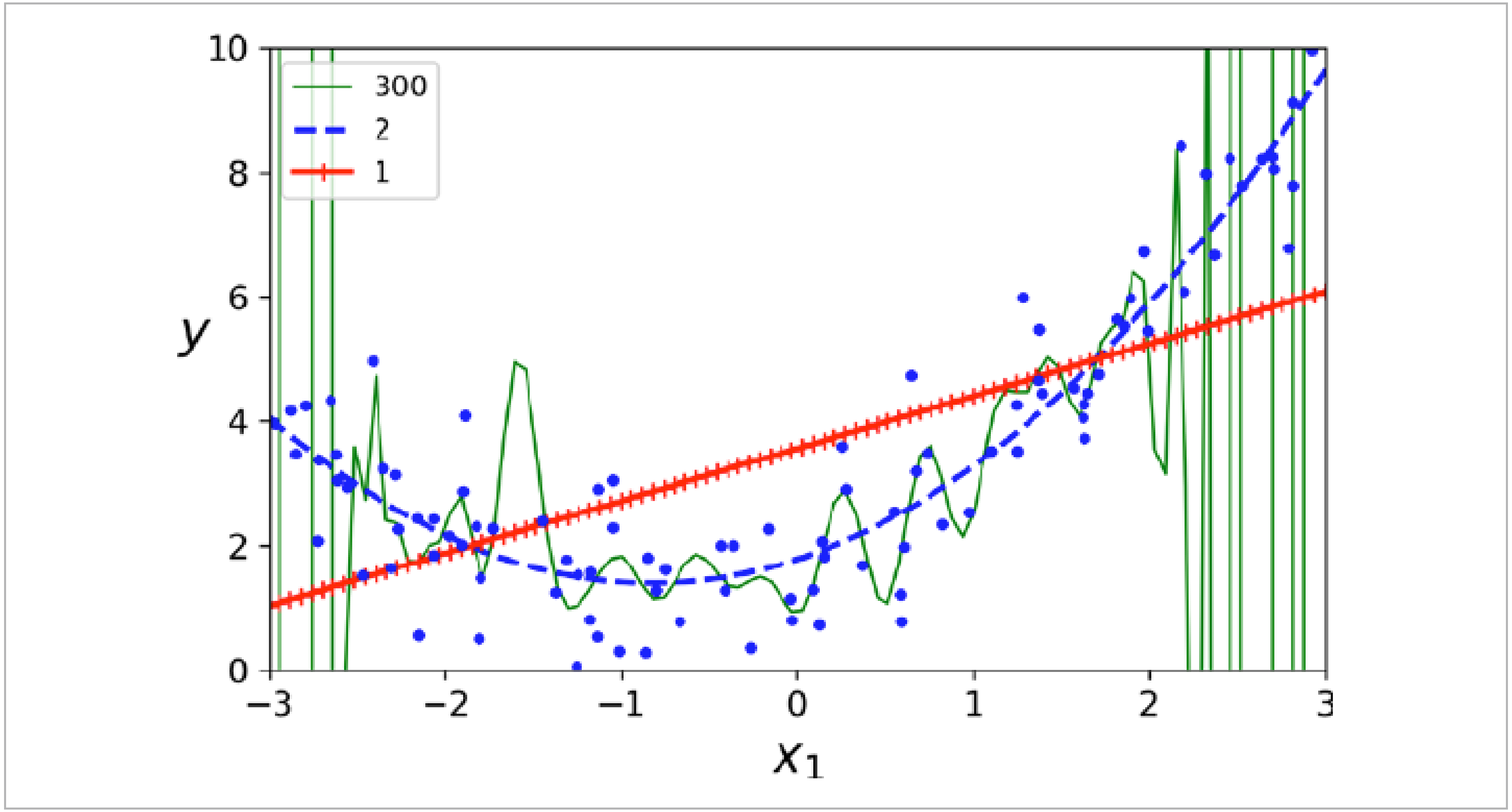
4-1. 선형 회귀

1. 정규방정식

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

- $\hat{\theta}$: 비용함수를 최소화하는 θ값
- y : y(1)부터 y(m)까지 포함하는 타깃 벡터

4-4. 학습 곡선



4-4. 학습 곡선

- 모델이 데이터에 과대/과소적합 되었는지 알아보는 방법
  - 교차 검증 → 2장에서 다루었음
  - **학습 곡선 살펴보기**
    - ✓ 훈련 세트와 검증 세트의 모델 성능을 훈련 세트 크기(또는 훈련 반복)의 함수로 나타낸다.
    - ✓ 단순히 훈련 세트에서 크기가 다른 서브 세트를 만들어 모델을 여러 번 훈련시키면 그래프를 생성할 수 있다.

4-4. 학습 곡선

- 단순 선형 회귀 모델(직선)의 학습 곡선

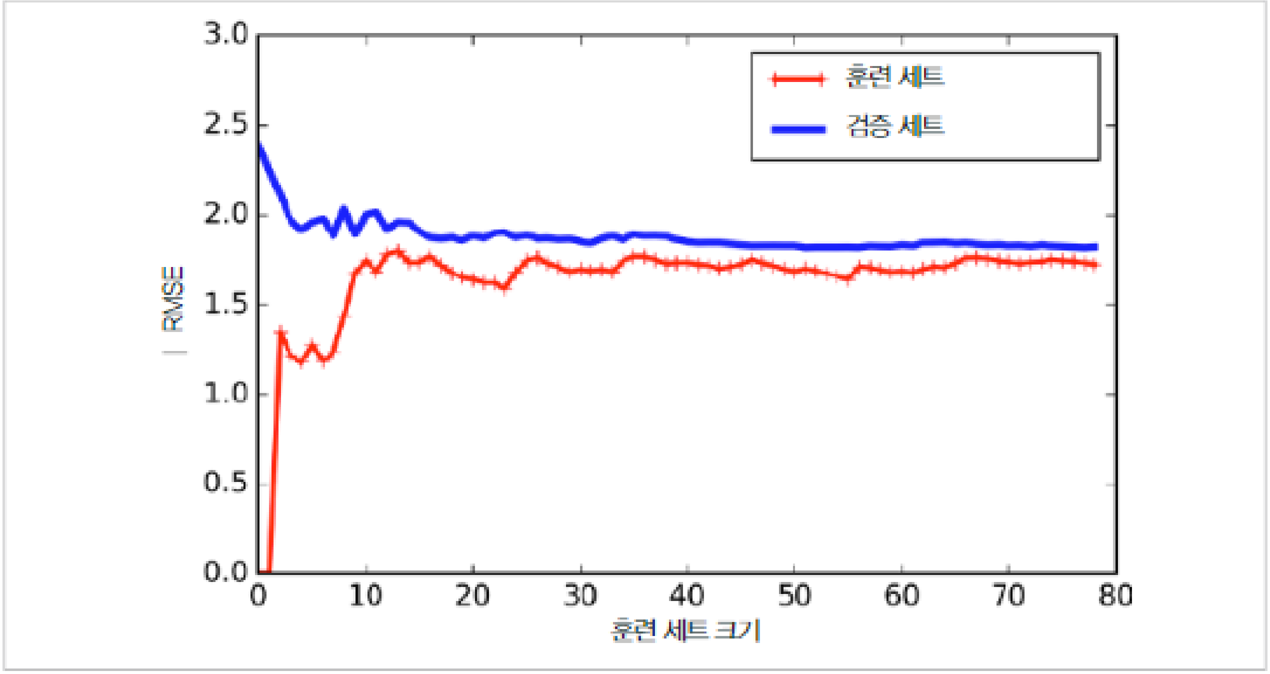


그림 4-15 학습 곡선

4-4. 학습 곡선

- 10차 다항 회귀 모델의 학습 곡선

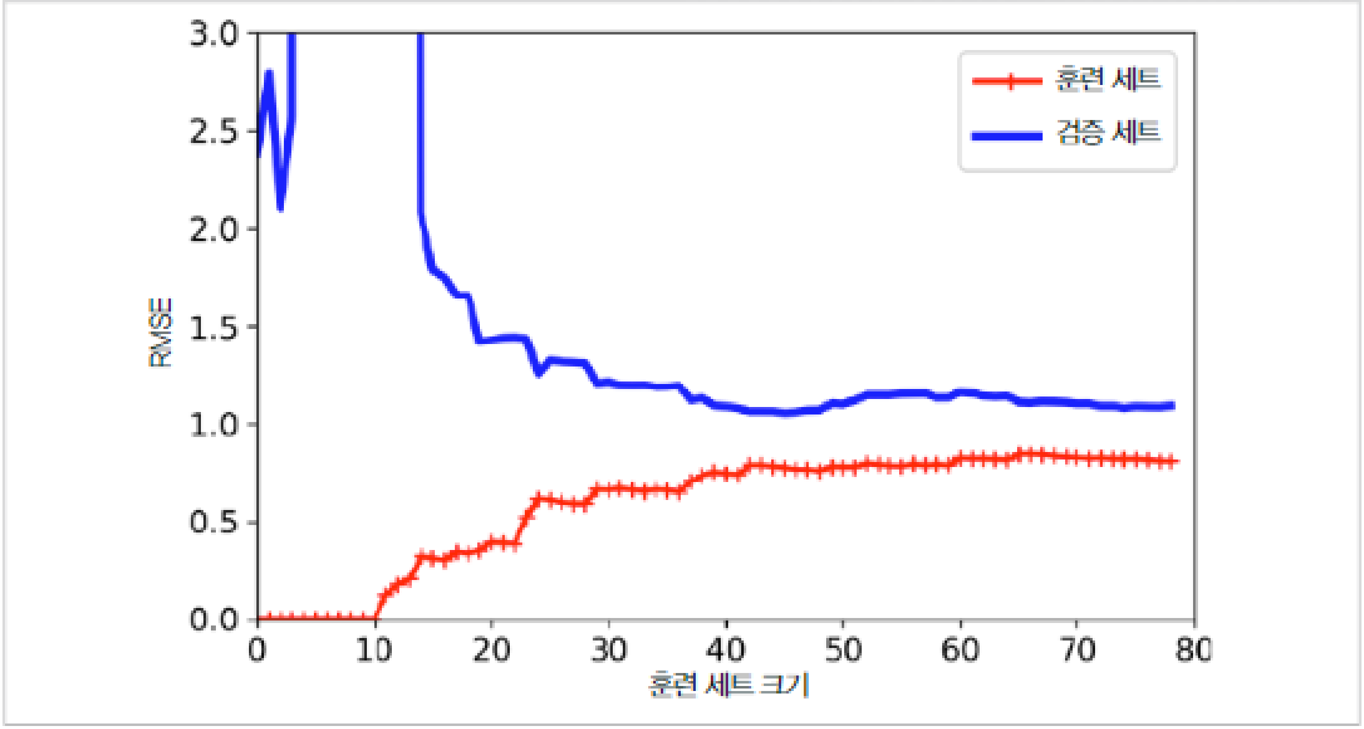


그림 4-16 10차 다항 회귀의 학습 곡선

4-5. 규제가 있는 선형 모델

- 과대적합 감소시키는 방법

  → **규제**

  ▪ 다항 회귀 모델

    ✓다항식의 차수 감소

  ▪ 선형 회귀 모델

    ✓가중치 제한

    ✓릿지 회귀, 라쏘 회귀, 엘리스틱넷

4-5.1. 릿지 회귀(티호노프 규제)

- 규제가 추가된 선형 회귀 버전
- $\alpha \sum_{i=1}^{n} {\theta_i}^2$ 이 비용 함수에 추가됨
  - 학습 알고리즘을 데이터에 맞추는 것뿐만 아니라 모델의 가중치가 가능한 작게 유지되도록 노력함
  - 규제항은 훈련하는 동안에만 비용 함수에 추가됨
  - 모델의 성능은 규제가 없는 성능 지표로 평가

4-5.1. 릿지 회귀(티호노프 규제)

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} {\theta_i}^2$$

- α : 모델을 얼마나 많이 규제할지 조절

  : 0이면 선형 회귀와 같아짐

  : 아주 크면 모든 가중치가 거의 0에 가까워지고 결국 데이터 평균을 지나는 수평선이 됨

4-5.1. 릿지 회귀(티호노프 규제)

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} {\theta_i}^2$$

- θ_0 : 규제되지 않음

  : 합 기호가 i=0이 아니고 i=1에서 시작하기 때문

4-5.1. 릿지 회귀(티호노프 규제)

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} {\theta_i}^2$$

- w를 특성의 가중치 벡터($\theta_1 \sim \theta_n$)라고 정의하면 규제항은 $\frac{1}{2}(||w||_2)^2$와 같음
  - $||*||_2$는 가중치 벡터의 norm
  - 경사 하강법에 적용하려면 MSE 그레디언트 벡터에 αw를 더하면 됨

4-5.1. 릿지 회귀(티호노프 규제)

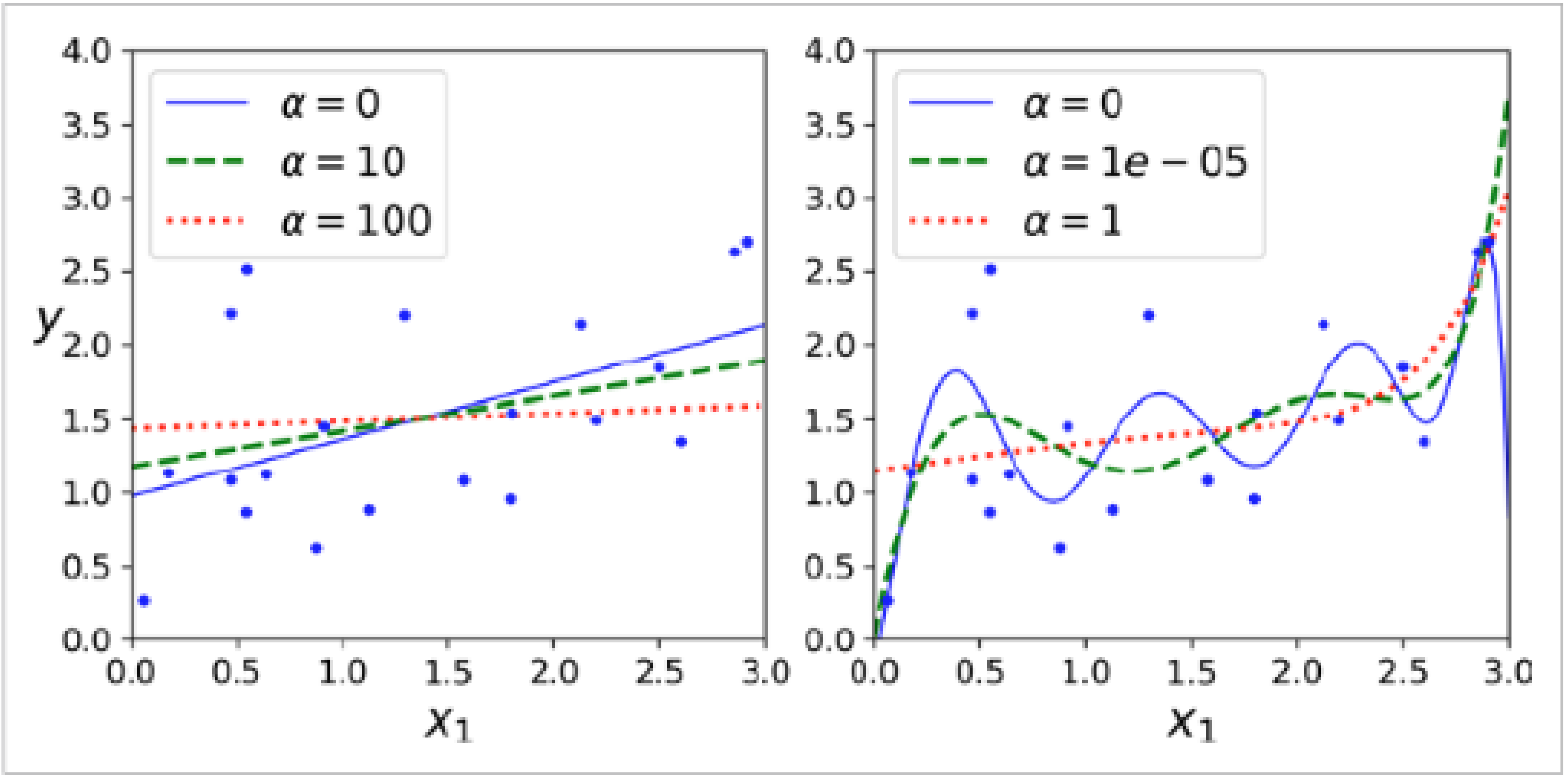


그림 4-17 다양한 수준의 릿지 규제를 사용한 선형 회귀(왼쪽)와 다항 회귀(오른쪽).

4-5.1. 릿지 회귀(티호노프 규제)

- 정규 방정식

$$\theta = (X^T X + \alpha A)^{-1} X^T y$$

  - A: 편향에 해당하는 맨 왼쪽 위의 원소가 0인 (n+1)*(n+1)의 단위행렬

4-5.2. 라쏘 회귀

- 선형 회귀의 또 다른 규제된 버전

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

4-5.2. 라쏘 회귀

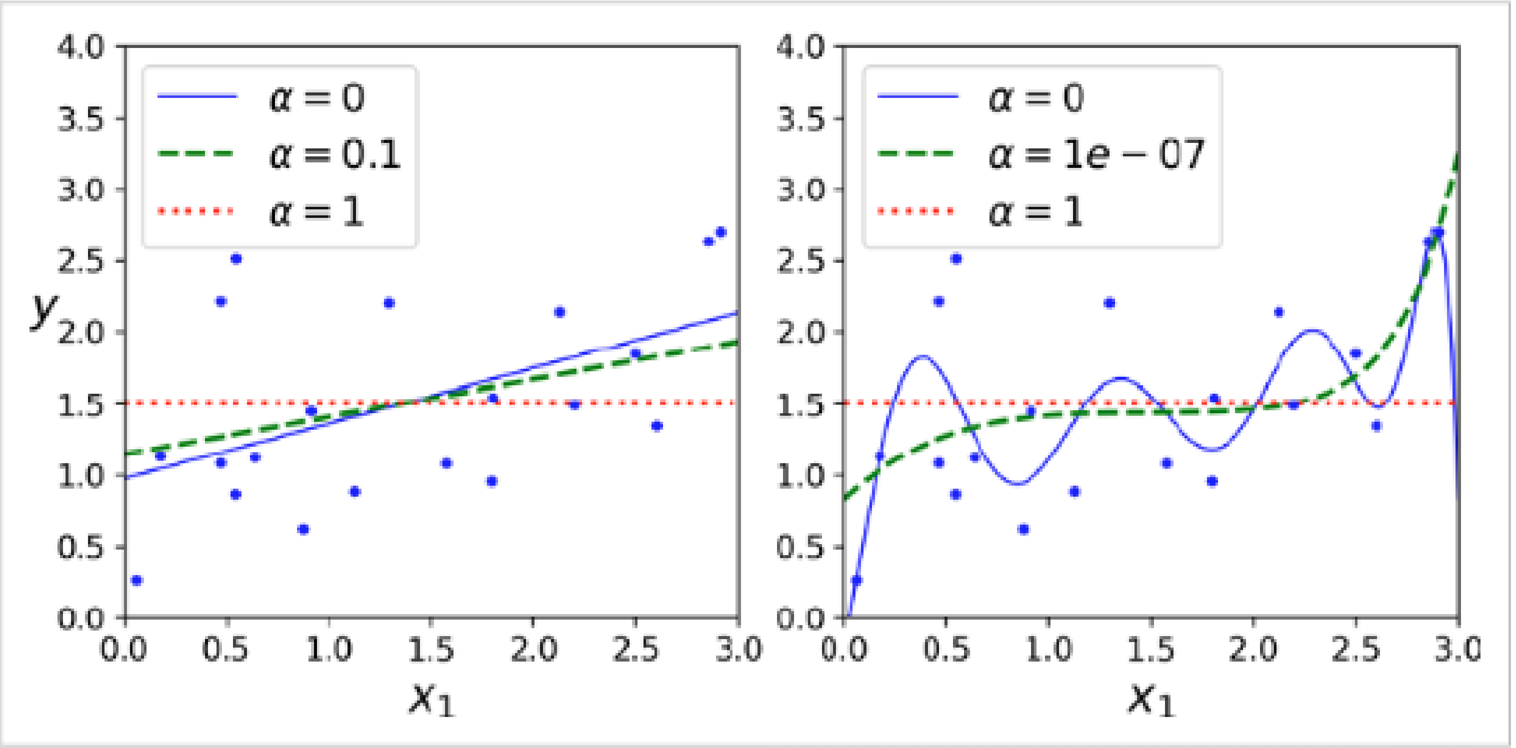


그림 4-18 다양한 수준의 라쏘 규제를 사용한 선형 회귀(왼쪽)와 다항 회귀(오른쪽).

4-5.2. 라쏘 회귀

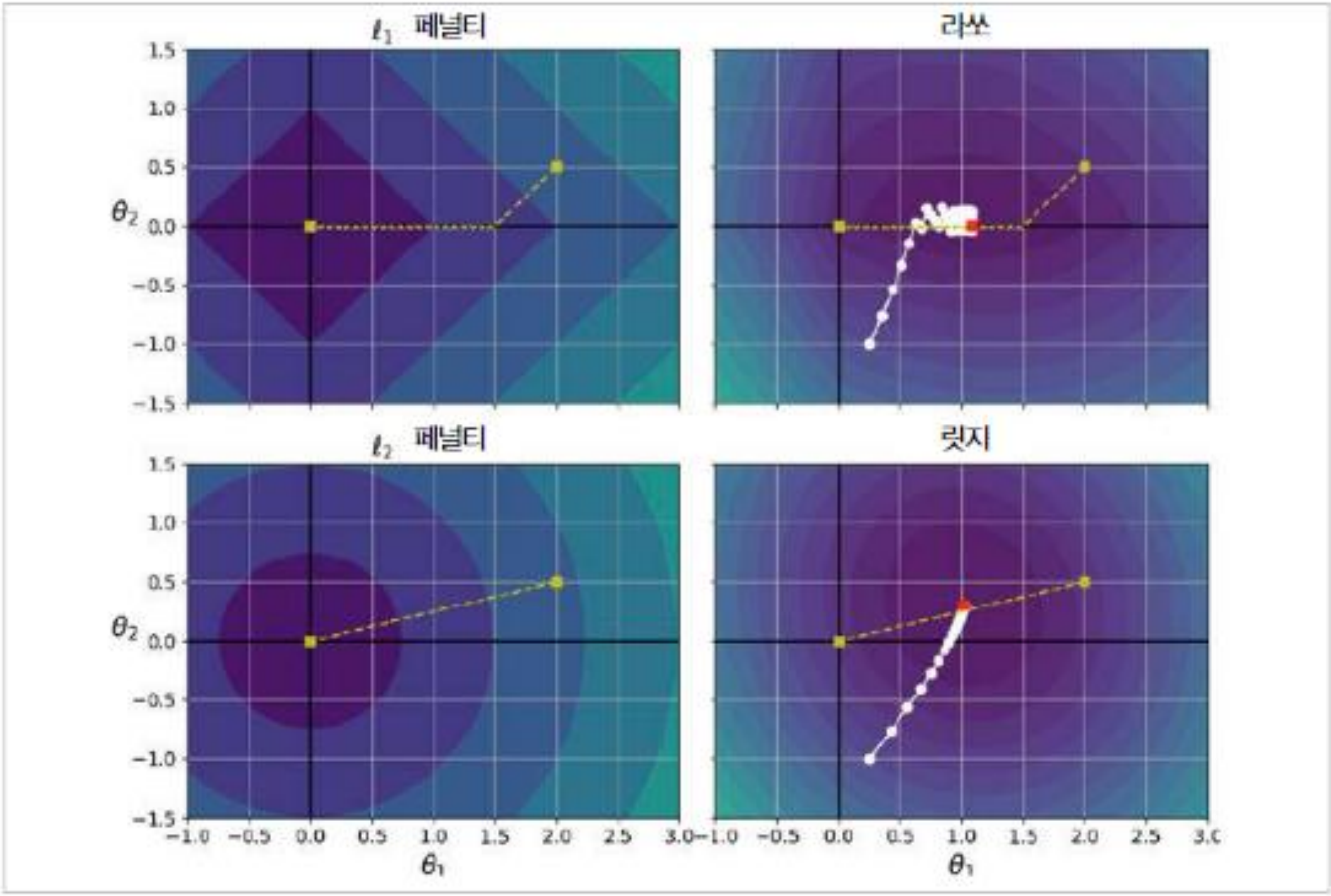


그림 4-19 라쏘 대 릿지 규제

4-5.2. 라쏘 회귀

$$g(\theta, J) = \nabla_\theta MSE(\theta) + \alpha \begin{pmatrix} sign(\theta_1) \\ \vdots \\ sign(\theta_n) \end{pmatrix}$$

$$sign(\theta_i) = \begin{cases} -1 & \theta_i < 0 \\ 0 & \theta_i = 0 \\ 1 & \theta_i > 0 \end{cases}$$

4-5.3. 엘라스틱넷

- 릿지 회귀와 라쏘 회귀를 절충한 모델
- 규제항은 릿지와 회귀의 규제항을 단순히 더해서 사용
- 혼합 정도는 혼합 비율 r을 사용해 조절
  - r=0이면 릿지 회귀
  - r=1이면 라쏘 회귀

$$J(\theta) = MSE(\theta) + r\alpha \sum_{i=1}^{n} |\theta_i| + \frac{r-1}{2}\alpha \sum_{i=1}^{n} {\theta_i}^2$$

4-5. 규제가 있는 선형 모델

선형 회귀

릿지

라쏘

엘라스틱넷

4-5.4. 조기 종료

- 검증 에러가 최솟값에 도달하면 바로 훈련을 중지시키는 것

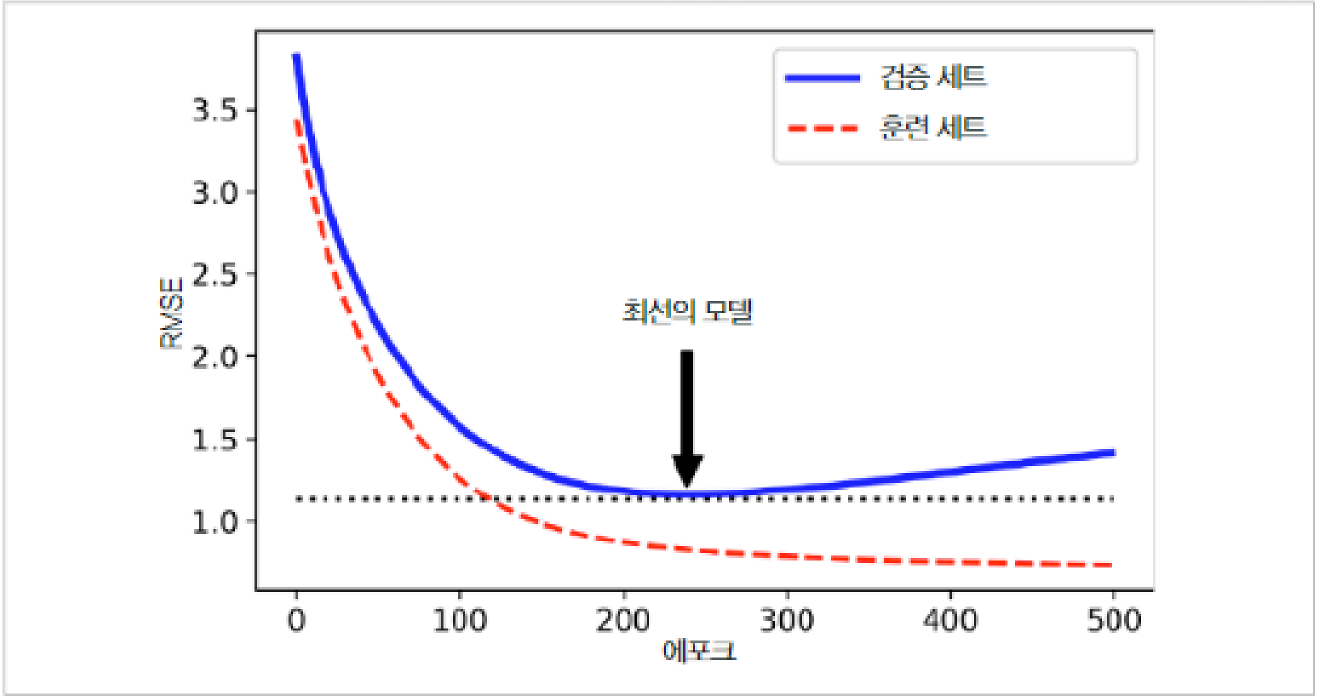


그림 4-20 조기 종료 규제