N / A / N / O / D / E / G / R / E / E

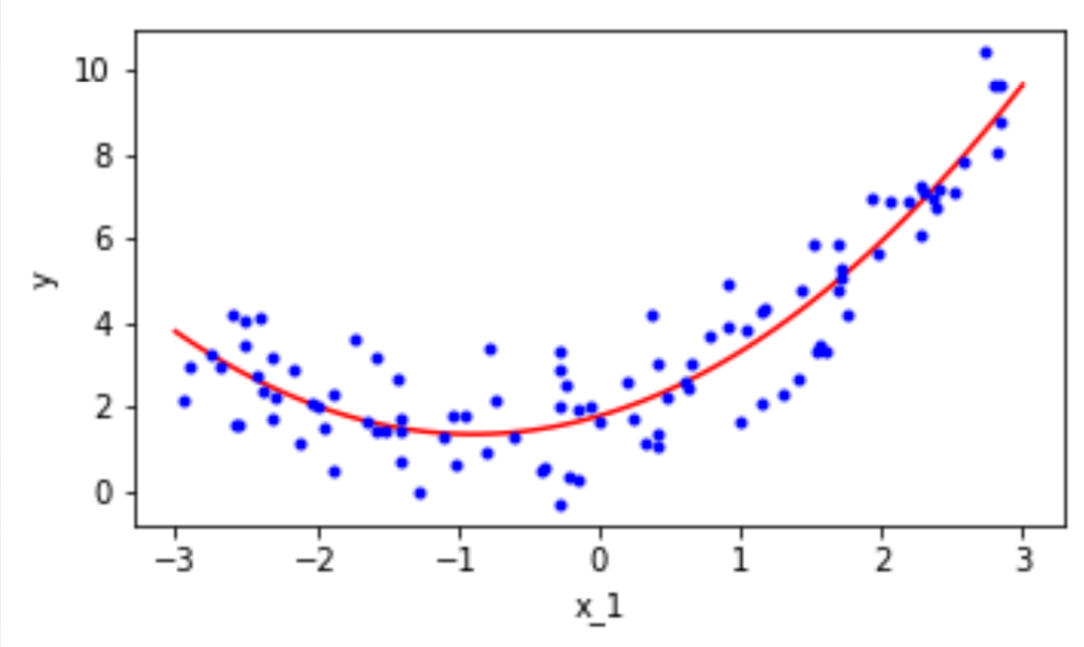
지도학습 알고리즘

04. 선형 회귀 III

Polynomial regression

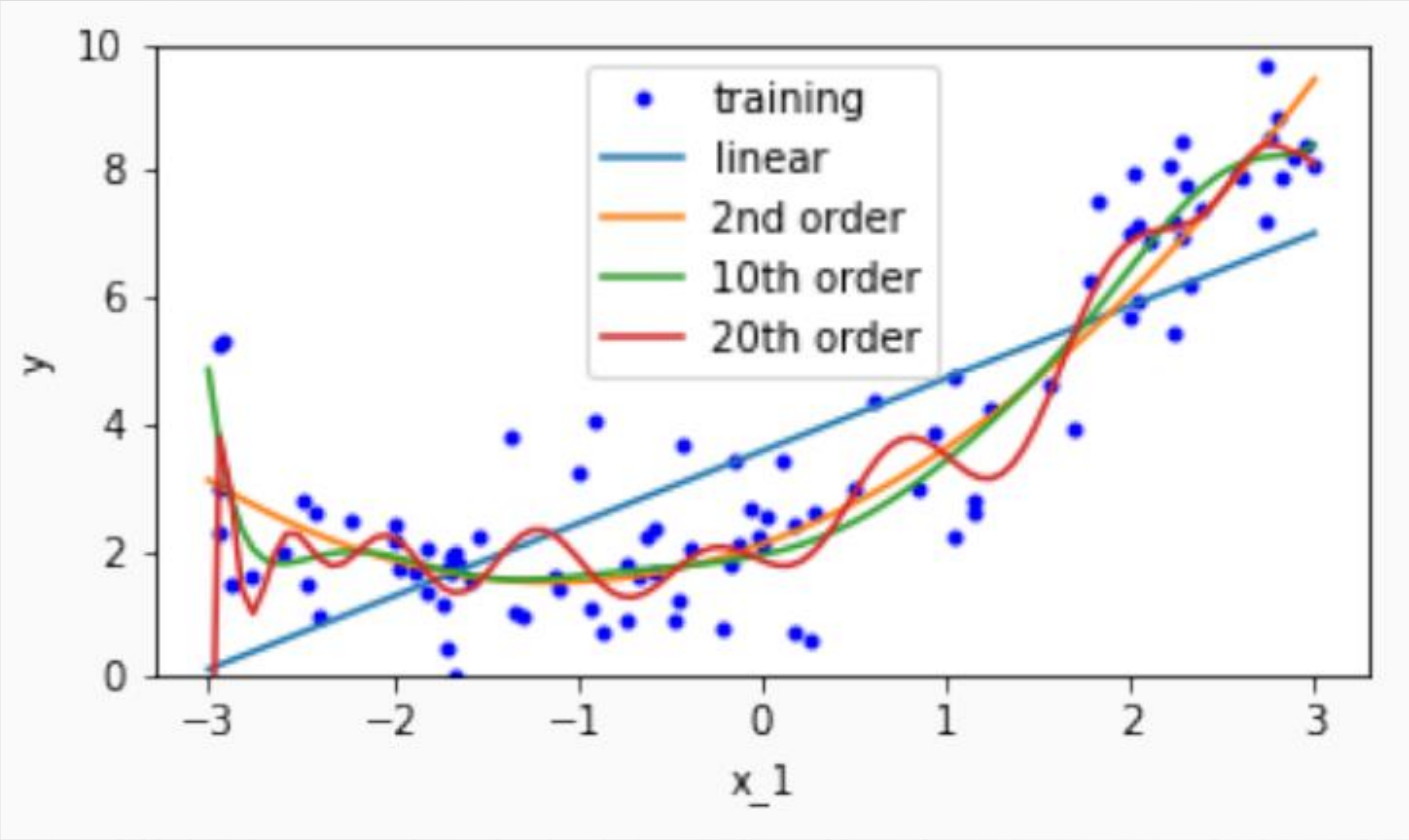
- 데이터들이 비선형 관계를 이루고 있을 때
- Polynomial의 각 항을 별개의 feature로 간주
- 동일하게 normal equation이나 gradient descent의 적용 가능

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$



Polynomial regression

- 모델의 차수는 어떻게 결정해야 하는가?

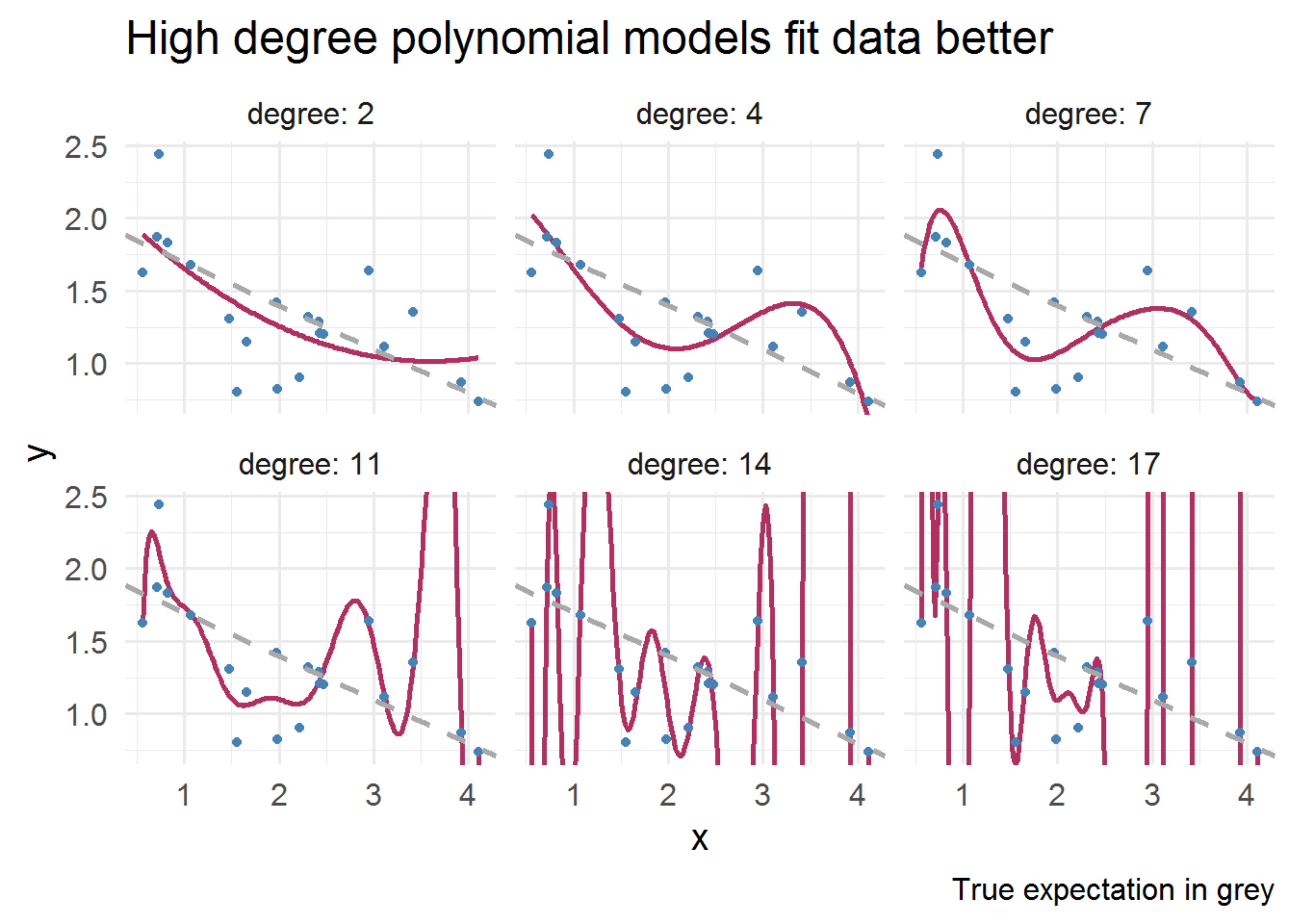


Overfitting and underfitting

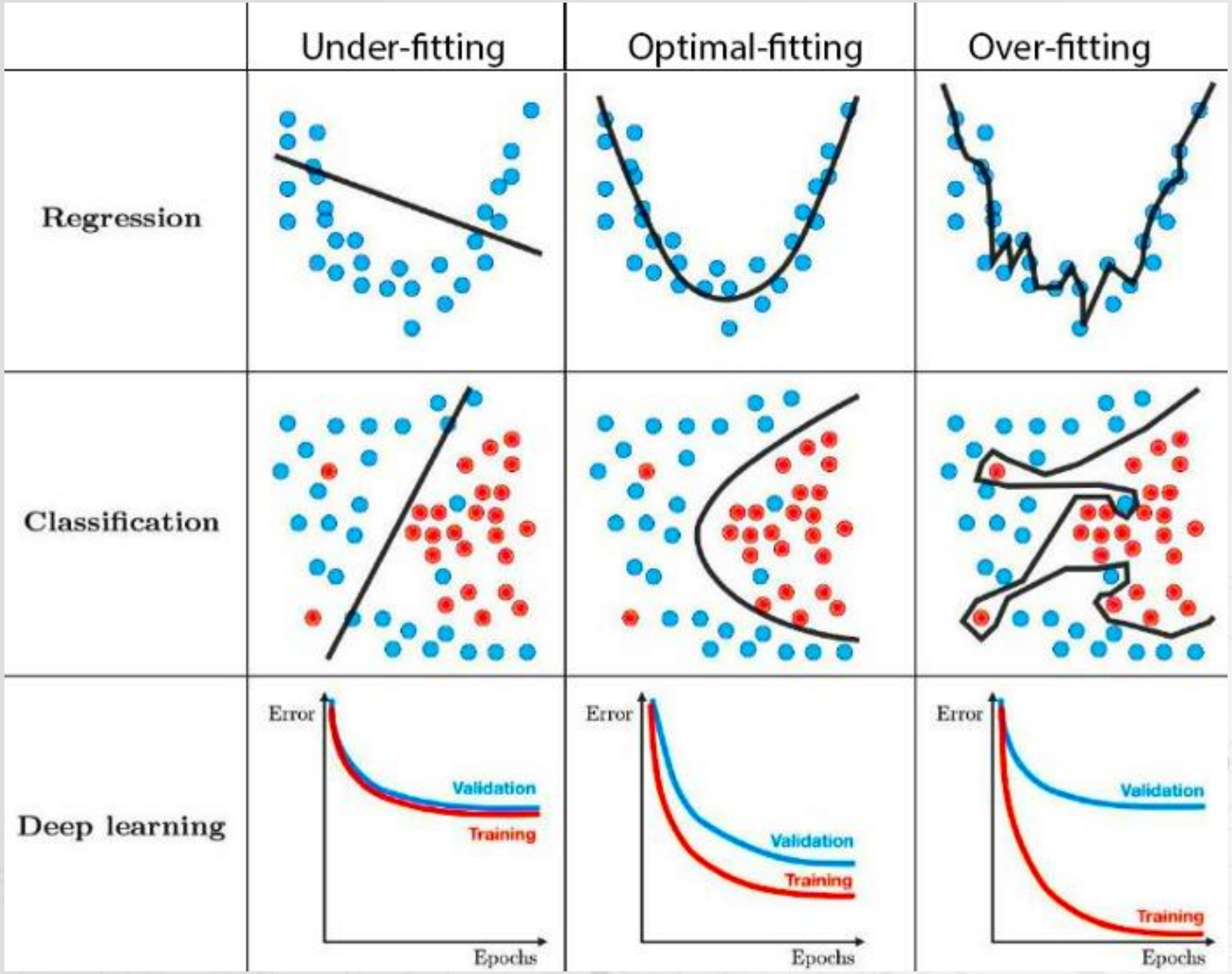
- 과적합 overfitting
  - 학습 데이터에 대해서는 좋은 성능을 보이지만, 처음 보는 데이터에 대해서는 일반화하지 못함
  - 학습 데이터의 양이나 노이즈의 정도에 비해 모델이 너무 복잡할 때
  - 모델이 학습 데이터를 일반화하는 범위를 넘어서 과도하게 의존함
  - 모델이 학습 데이터를 기억하기 시작함

- 미적합 underfitting
  - 모델이 너무 단순하여 학습 데이터에 대해서도 부정확

Overfitting and underfitting



Overfitting and underfitting



Ridge regression

- 비용 함수에 model parameter에 대한 penalty를 추가
- 과적합을 방지하면서 동시에 model parameter도 가능한 작게 만들도록

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \frac{\alpha}{2}\|\boldsymbol{\theta}\|^2$$

- Hyperparameter α를 통해 상대적인 비중을 조정
- Closed form solution

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X} + \alpha\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{y}$$

Ridge regression

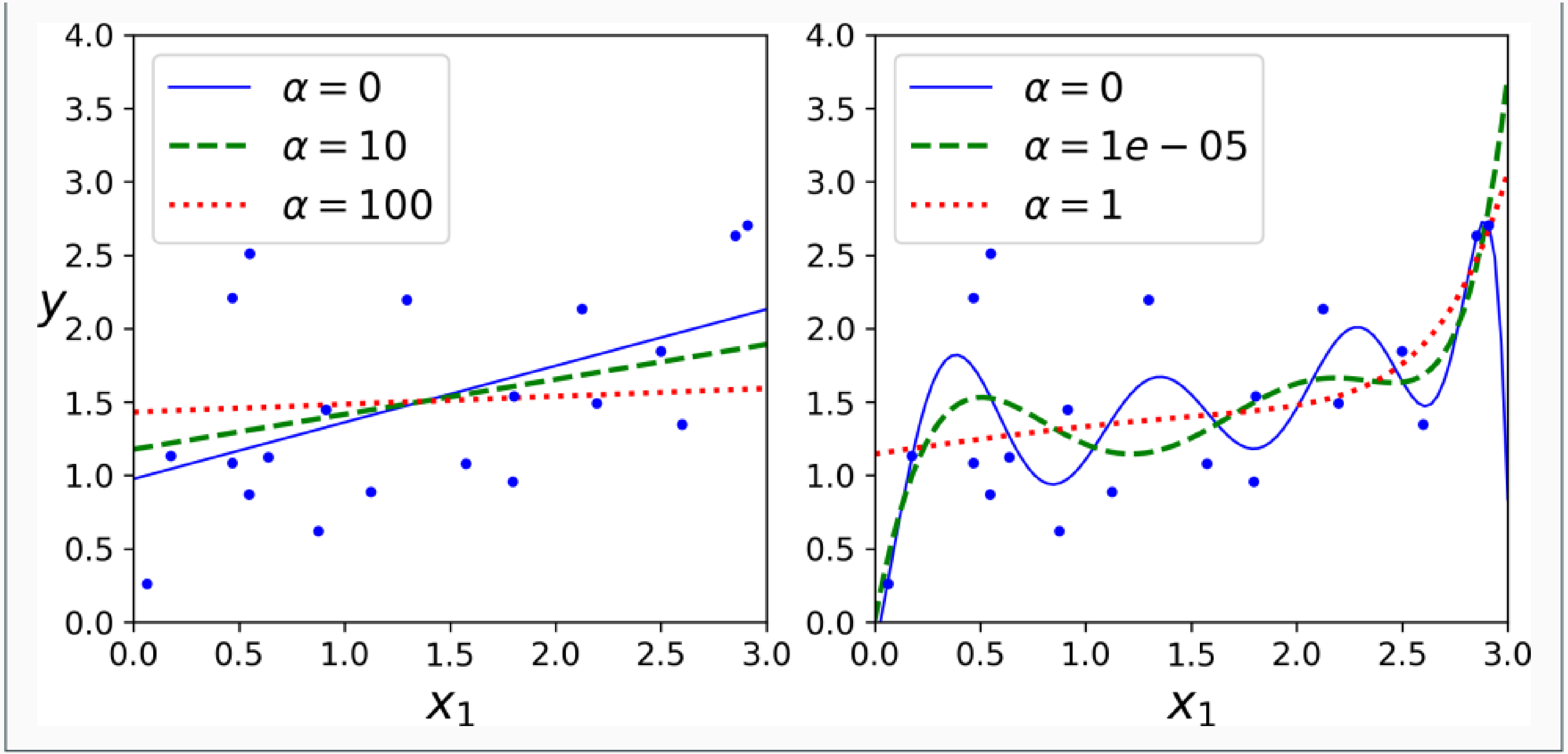


Figure 4-17. Ridge Regression

출처: A. Geron, Hands-On Machine Learning with Scikit-Learn and Tensorflow

Lasso regression

- Least absolute shrinkage and selection operator regression (LASSO)

- 중요하지 않은 feature에 대해서는 parameter가 0으로 나오는 경향
- Feature selection에 이용이 가능함

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \alpha \sum_i |\theta_i|$$

Lasso regression

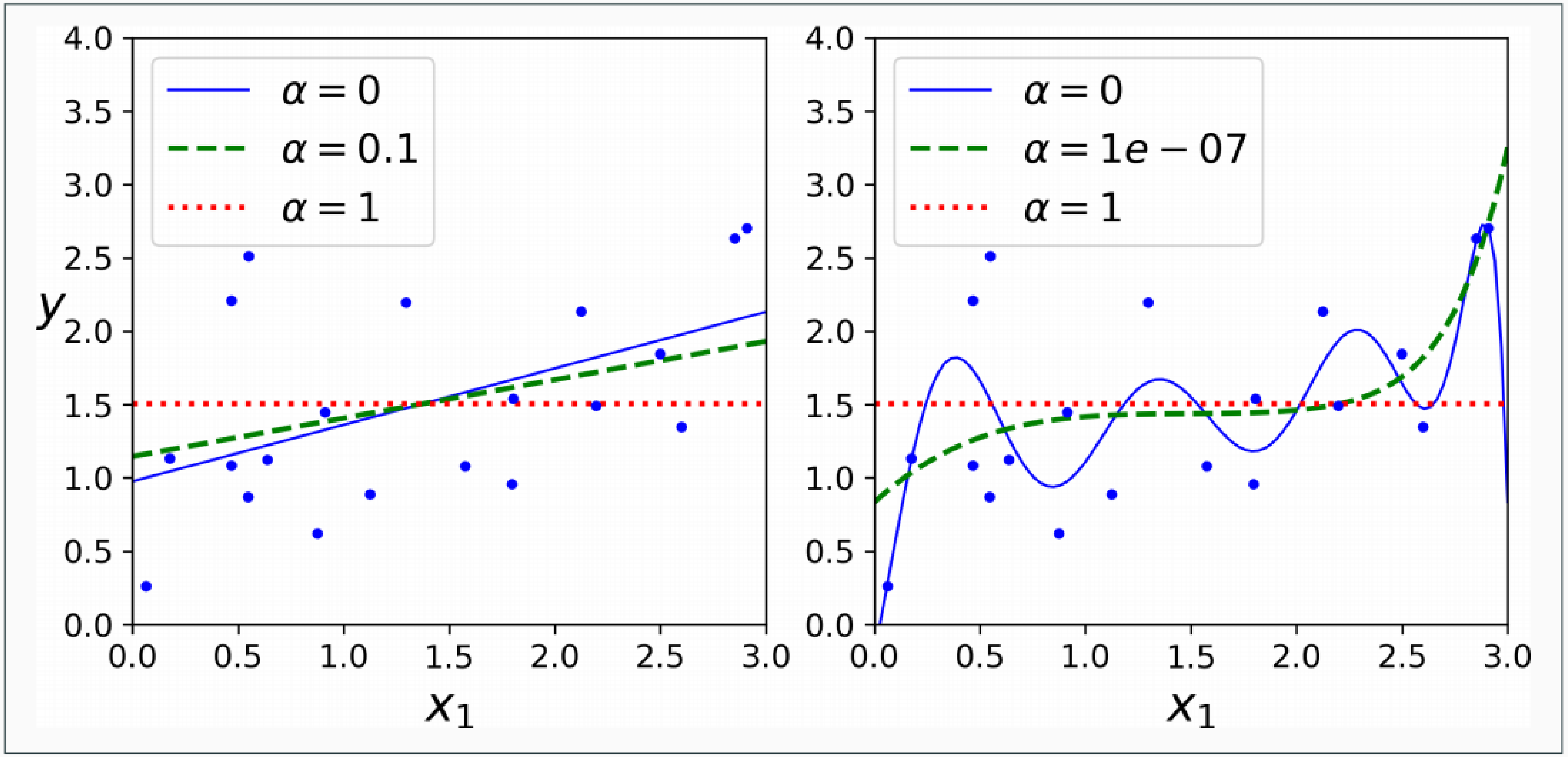


Figure 4-18. Lasso Regression

출처: A. Geron, Hands-On Machine Learning with Scikit-Learn and Tensorflow