Inclass 19: Polynomial Regression and Regularization

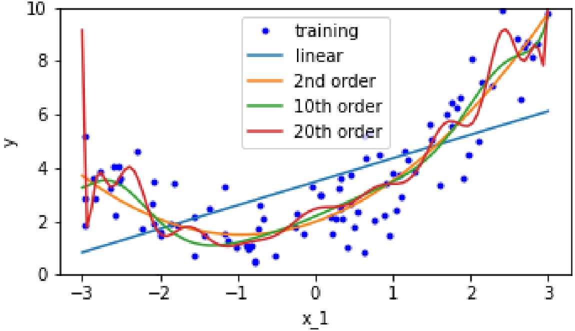
[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

Polynomial regression

모델의 차수는 어떻게 결정해야 하는가?



Overfitting and underfitting

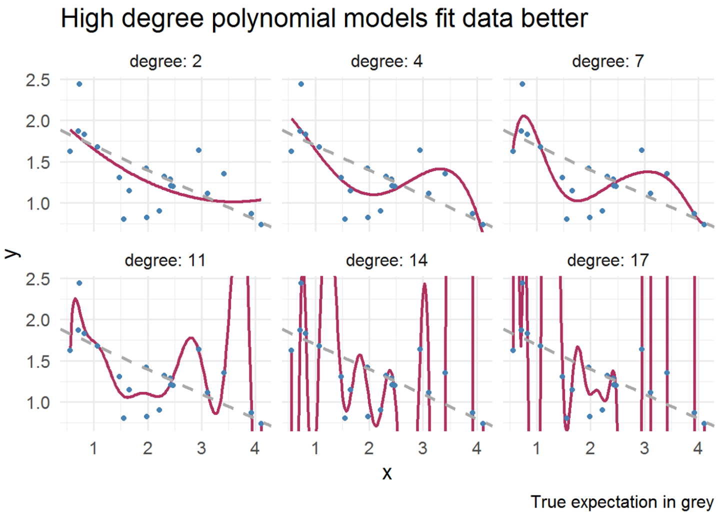
과적합 overfitting

- 학습 데이터에 대해서는 좋은 성능을 보이지만
- 처음 보는 데이터에 대해서는 일반화하지 못함
- 학습 데이터의 양이나 노이즈에 비해 모델이 너무 복잡할 때
- 모델이 학습 데이터를 일반화하는 범위를 넘어서 과도하게 의존함
- 모델이 학습 데이터를 기억하는듯한 양상을 보이기 시작

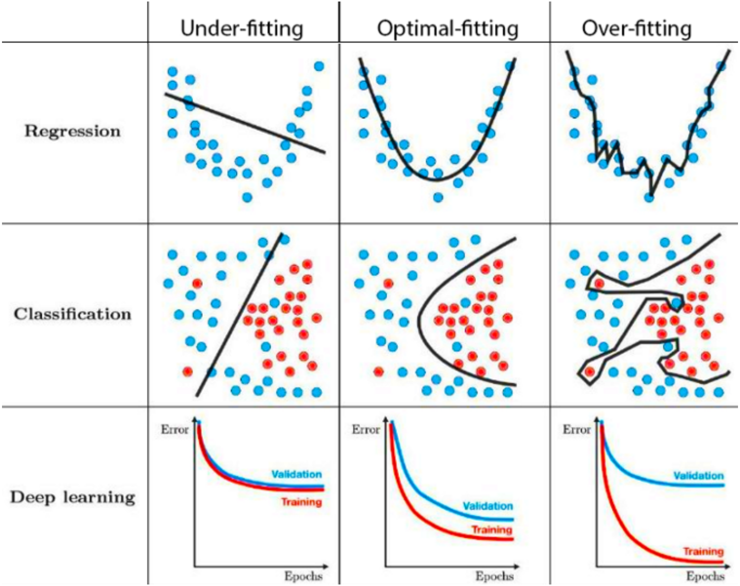
미적합 underfitting

- 모델이 너무 단순하여 학습 데이터에 대해서도 부정확

Overfitting and underfitting



Overfitting and underfitting



Regularization

정규화 regularization

- 비용 함수에 복잡도에 대한 penalty를 추가
- 과적합에 대한 비용을 optimizing cost에 반영

Ridge regression

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \frac{\alpha}{2}\|\boldsymbol{\theta}\|^2 \tag{1}$$

LASSO

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \alpha \sum_i |\theta_i| \tag{2}$$

Ridge regression

Ridge regression

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \frac{\alpha}{2}\|\boldsymbol{\theta}\|^2 \tag{3}$$

- Hyperparameter α로 두 항목 간의 상대적인 비중을 조절
- Closed form solution $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X} + \alpha\mathbf{I}_n)^{-1}\mathbf{X}^T\mathbf{y}$

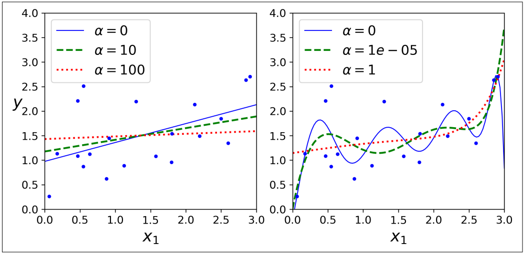


Figure 4-17. Ridge Regression

LASSO

$$J(\boldsymbol{\theta}) = \mathrm{SSE}(\boldsymbol{\theta}) + \alpha \sum_{i=1}^{n} |\theta_i| \tag{4}$$

- Least Absolute Shrinkage and Selection Operator Regression
- 중요하지 않은 feature들의 weight를 0으로 만드는 경향

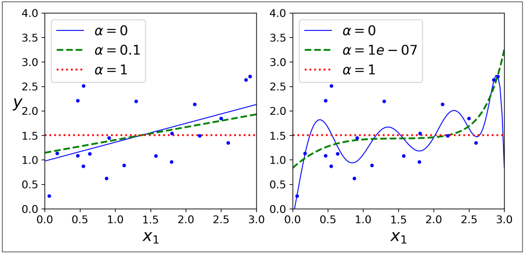


Figure 4-18. Lasso Regression