

Deep Learning

2.4 Regularization

Dr. Konda Reddy Mopuri
kmopuri@iittp.ac.in
Dept. of CSE, IIT Tirupati

Regularization

- ① A central problem in ML: Generalization

Regularization

- ① A central problem in ML: Generalization
- ② Many strategies are explicitly designed to reduce the test error at the expense of training error

Regularization

- ① A central problem in ML: Generalization
- ② Many strategies are explicitly designed to reduce the test error at the expense of training error
- ③ These strategies collectively are called 'Regularization'

Regularization

- ① A central problem in ML: Generalization
- ② Many strategies are explicitly designed to reduce the test error at the expense of training error
- ③ These strategies collectively are called 'Regularization'
- ④ Developing effective regularization techniques has been a major research interest in the field

Polynomial for regression problem



- ① Modifies the model's representational capacity by adding or removing functions from the hypothesis space

Polynomial for regression problem

- ① Modifies the model's representational capacity by adding or removing functions from the hypothesis space
- ② Algorithm's behaviour is strongly affected by not only the size of the hypothesis space, but also the specific identity of the functions

Regularization

- ① Preferring some functions over others in the hypothesis space

Regularization

- ① Preferring some functions over others in the hypothesis space
- ② Unpreferred solution will be picked if it fits the training data significantly better than the preferred one

Linear Regression with weight decay

- ① Include weight decay in the training criterion

$$\mathcal{L}(w) = MSE_{train} + \lambda w^T w$$

Linear Regression with weight decay

- 1 Include weight decay in the training criterion

$$\mathcal{L}(w) = MSE_{train} + \lambda w^T w$$

- 2 Prefers solutions with smaller l_2 norm of the weights (smaller slope or put weight on fewer features)

Linear Regression with weight decay

- ① True function is quadratic

Linear Regression with weight decay

- ① True function is quadratic
- ② We fit models with degree 9

Linear Regression with weight decay

- ① True function is quadratic
- ② We fit models with degree 9
- ③ Vary the amount of weight decay to prevent overfitting

Linear Regression with weight decay

- ① True function is quadratic
- ② We fit models with degree 9
- ③ Vary the amount of weight decay to prevent overfitting

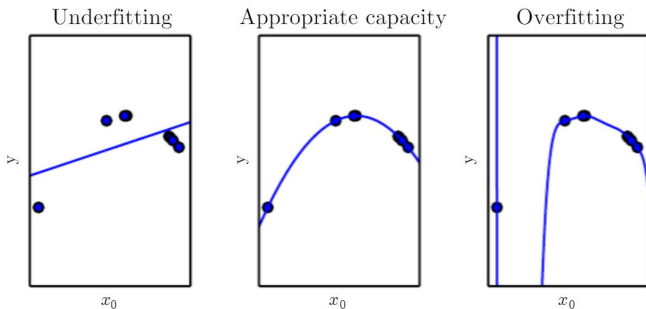


Figure from deeplearningbook.org

Regularization

- ① The penalty term added to the cost is called a 'regularizer'

Regularization

- ① The penalty term added to the cost is called a 'regularizer'
- ② Expressing preferences is a general way of controlling model capacity than inclusion or exclusion of functions

Regularization

- ① Any modification that we make to the learning algorithm with an intention to reduce the generalization error

Regularization

- ① Any modification that we make to the learning algorithm with an intention to reduce the generalization error
- ② No best regularization, need to pick the one suitable