ECE595 / STAT598: Machine Learning I

Lecture 31.1: Regularization - Motivation for Regularization

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE
UNIVERSITY

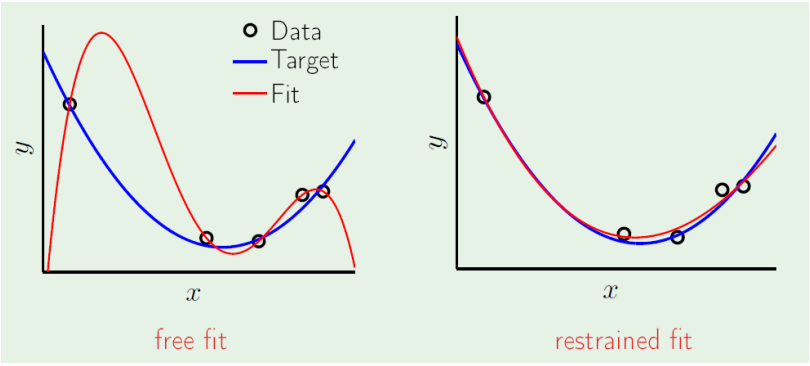
Outline

- Lecture 30 Overfit
- Lecture 31 Regularization
- Lecture 32 Validation

Today's Lecture:

- Motivation for Regularization
  - VC Analysis
  - Example
- Two Regularization Techniques
  - Weight Decay
  - Augmented Error
- Choosing a Regularization
  - Pill or Poisson?
  - Role of λ

Overcoming Overfit



- Regularization is one weapon to combat overfitting.
- Constrains the learning algorithm to improve out-sample error when noise is present.
- Regularization is as much an art as it is a science.

Regularization from VC Analysis

$$E_{\text{out}}(h) \leq E_{\text{in}}(h) + \Omega(\mathcal{H}), \qquad \text{for all } h \in \mathcal{H} \tag{1}$$

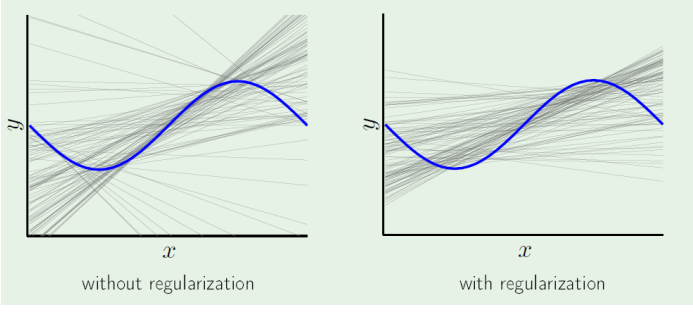
- Model complexity penalty $\Omega(\mathcal{H})$
- If you want $E_{\text{out}}(h)$ to be small, better to make $\Omega(\mathcal{H})$ small
- Roughly speaking, fit data using a "simple" h from $\mathcal{H}$
- So you are effectively minimizing

$$\underset{h}{\text{minimize}} \quad E_{\text{in}}(h) + \Omega(h),$$

- That is, instead of minimizing $E_{\text{in}}(h)$ only, you minimize $\Omega(h)$ too

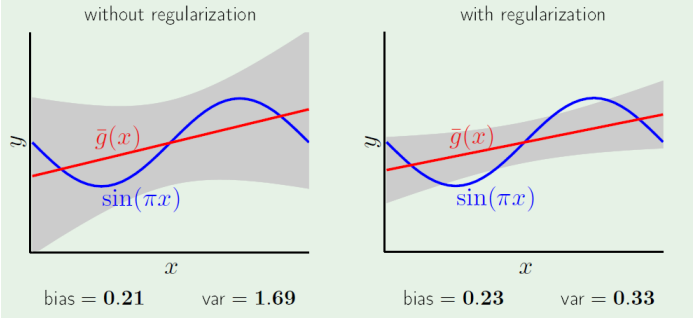
Example

- One regularization technique is **weight decay**.
- Measures the complexity of a hypothesis h by the size of the coefficients used to represent h (e.g., in a linear model).
- This technique prefers mild lines with small offset and slope.
- Applying this concept to the sine example before, trying to fit $N = 2$ data points, using $\mathcal{H}_1$ (the set of lines).



Example

- Recall the constant model: Fit the two data points using a constant line.
- Constant model has $E_{\text{out}} = 0.75$
- Unregularized model has $E_{\text{out}} = 1.90$
- Regularized model has $E_{\text{out}} = 0.56$
- Bias-variance: Improve variance but suffer from bias. Overall is better.



Why Need Regularization?

- The linear model is too sophisticated for the amount of data we have.
- A line can fit any two points!
- This problem is still here even if we change the target function.
- The need of regularization depends on **quantity** of data, and **quality** of data.
- Given only two points, we can either choose
  - a simple model, e.g., constant model
  - to constrain the model, e.g., weight decay
- Constraining the model gives us more flexibility.