ECE595 / STAT598: Machine Learning I

Lecture 30.1: Overfit - Source of Overfit

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE
UNIVERSITY

Outline

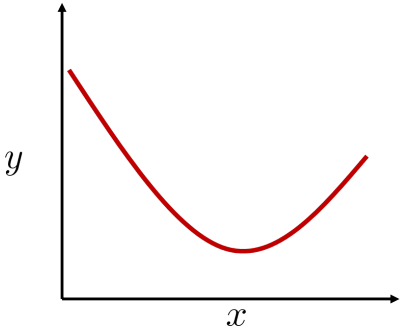
- Lecture 30 Overfit
- Lecture 31 Regularization
- Lecture 32 Validation

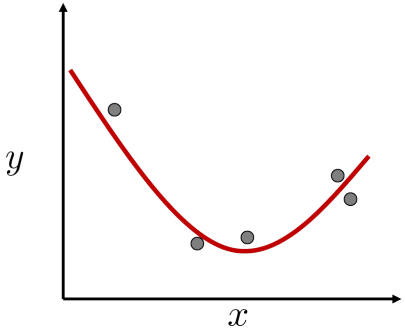
Today's Lecture:

- Source of Overfit
  - Is Noise the Reason?
  - Is Model Complexity the Reason?
  - The Trinity of Noise, Target Complexity, and Training Sample
- Analyzing Overfit
  - Bias and Variance
  - Learning Curve

Case Study

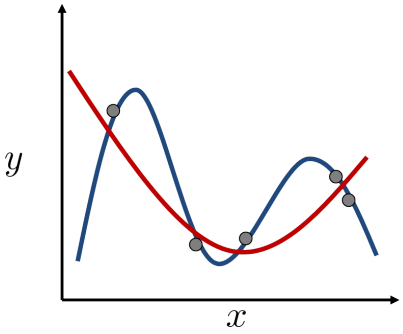
- What is overfit?
- You have a simple target function f
- From this f you generate 5 data **noisy** training samples
- Then you use a 4-th order polynomial to fit the data



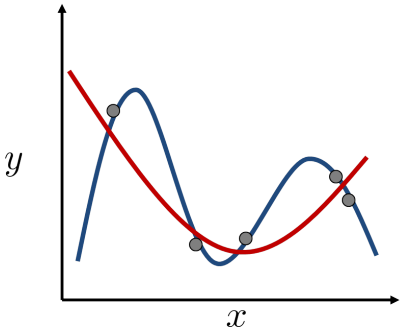


Case Study

- What is overfit?
- You have a simple target function f
- From this f you generate 5 data **noisy** training samples
- Then you use a 4-th order polynomial to fit the data



What is Over-fitting?

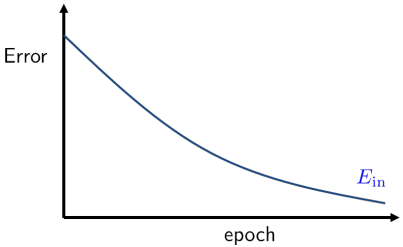


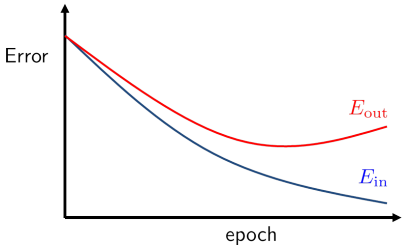
- If you use a 4-th order polynomial to fit the data
- $E_{\text{in}} = 0$: You have a perfect fit
- $E_{\text{out}} =$ terrible. You cannot generalize
- What could go wrong?

Is Model Complexity the Reason?

Common belief: Complex models tends to overfit. Is this true?

- Fix a neural network structure
- Fix training data
- Run for more epoches
- E_{in} drops
- E_{out} drops and then rises
- When you use more epoches, you fit the noise
- In this example, the network capacity is not changed, but you still have overfit

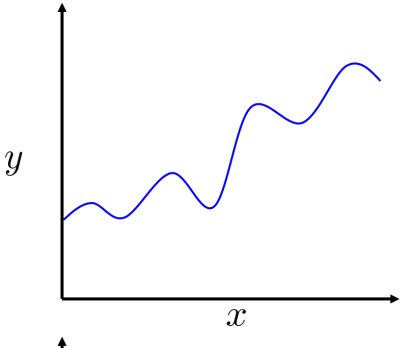


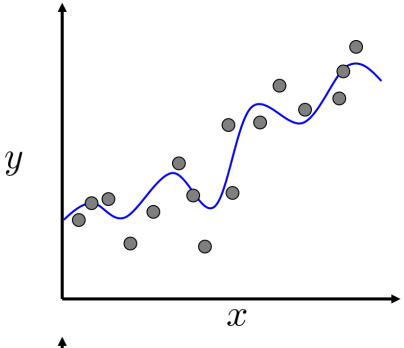


Is Noise the Reason?

Common belief: There is always noise, and so you overfit. Is it true?

- You have a 10-th order target function
- You generate N **noisy** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial

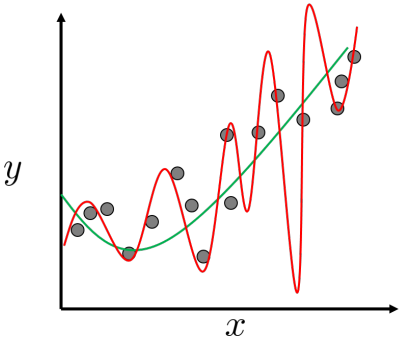




Is Noise the Reason?

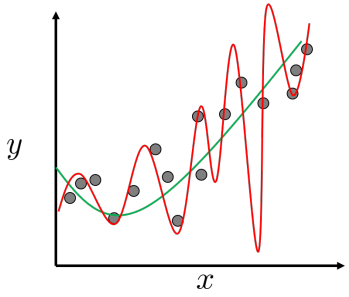
Common belief: There is always noise, and so you overfit. Is it true?

- You have a 10-th order target function
- You generate N **noisy** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial



Is Noise the Reason?

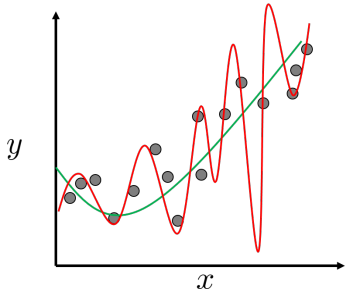
Let's look at the training and testing error.



- You have a 10-th order target function
- You generate N **noisy** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial: $E_{\text{in}} = 0.05$
  - (ii) Use a 10-th order polynomial: $E_{\text{in}} = 0.034$

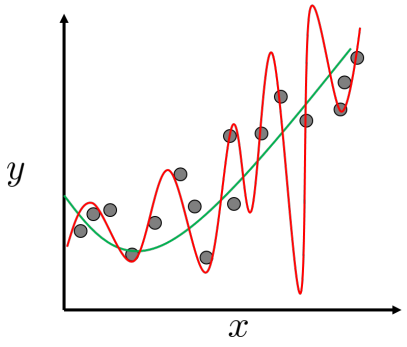
Is Noise the Reason?

Let's look at the training and testing error.



- You have a 10-th order target function
- You generate N **noisy** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial: $E_{\text{in}} = 0.05$, $E_{\text{out}} = 0.127$
  - (ii) Use a 10-th order polynomial: $E_{\text{in}} = 0.034$, $E_{\text{out}} = 9.00$

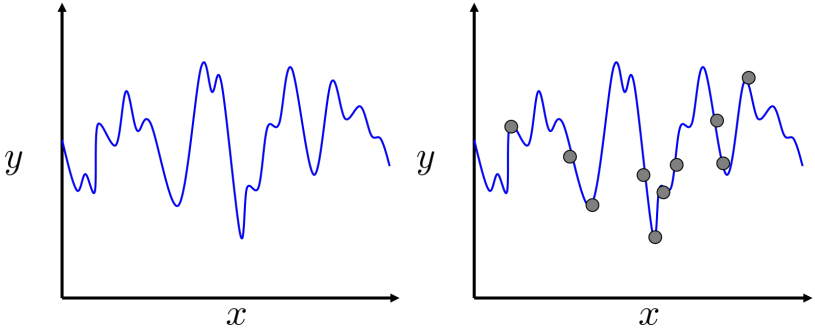
Is Noise the Reason?



- Noise is indeed a reason for overfitting
- You are fitting noise if you use complex models
- Trade-off: noise and model complexity
- But not the only reason

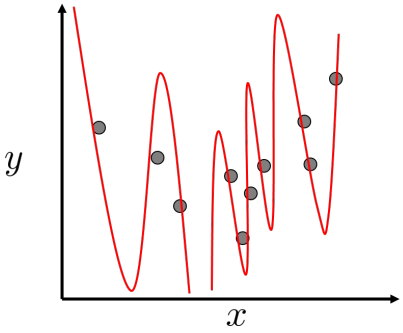
How about we Consider a Clean Target?

- You have a 50-th order target function
- You generate N **clean** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial

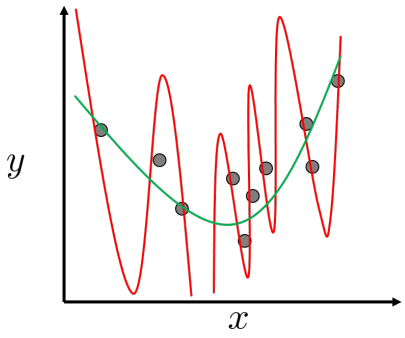


How about we Consider a Clean Target?

- You have a 50-th order target function
- You generate N **clean** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial

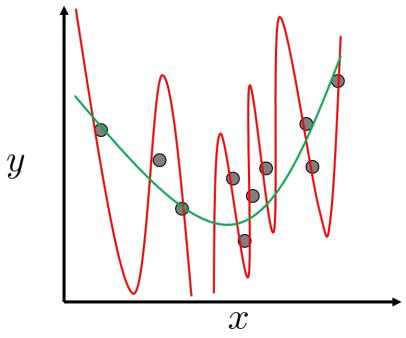


How about we Consider a Clean Target?



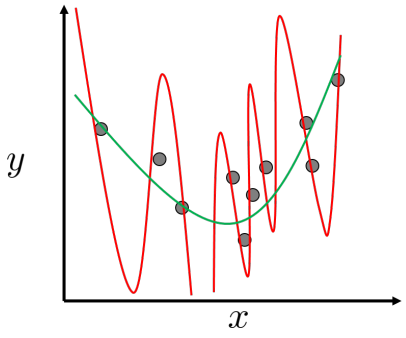
- You have a 50-th order target function
- You generate N **clean** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial: $E_{\text{in}} = 0.029$
  - (ii) Use a 10-th order polynomial: $E_{\text{in}} = 10^{-5}$

How about we Consider a Clean Target?



- You have a 50-th order target function
- You generate N **clean** observations
- How well can you fit if you
  - (i) Use a 2nd order polynomial: $E_{\text{in}} = 0.029$, $E_{\text{out}} = 0.120$
  - (ii) Use a 10-th order polynomial: $E_{\text{in}} = 10^{-5}$, $E_{\text{out}} = 7680$

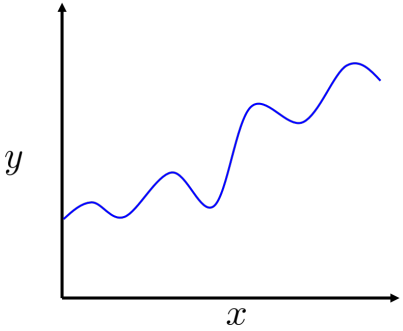
Is Noise the Reason?

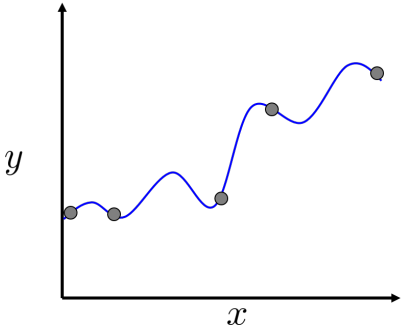


- Noise-free
- Still overfit
- Problem 1: Model mismatch? 50th order target and 10th order fit?
- Problem 2: Lack of training samples?

Is Model Fit a Reason?

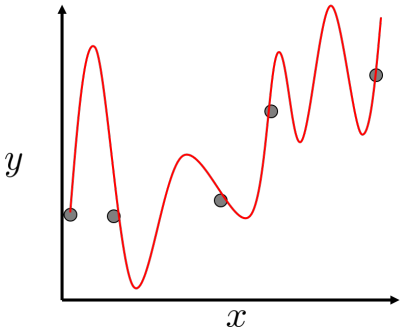
- What if we *know* that the target function is 10-th order
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial
- The samples are clean



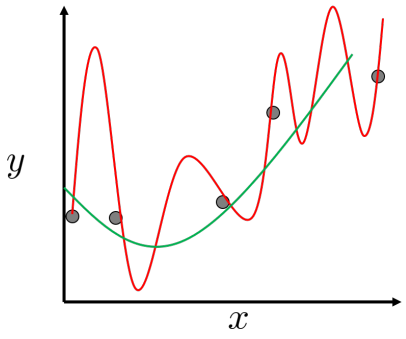


Is Model Fit a Reason?

- What if we *know* that the target function is 10-th order
- How well can you fit if you
  - (i) Use a 2nd order polynomial
  - (ii) Use a 10-th order polynomial
- The samples are clean



Is Model Fit a Reason?



- Noise-free
- Model match
- Does not fit well
- Problem: Lack of training samples

What Causes Overfit?

It is the trinity of

- **Training samples**
  - Not enough training samples?
  - Impossible to fit a complex model
  - Impossible to deal with noise
- **Target complexity**
  - Very complex target?
  - Need a lot of training samples to fit well
  - Not enough training samples, then can only use less complex hypotheses
- **Noise**
  - A lot of noise?
  - Need more training samples to average out the effect of noise