# ECE595 / STAT598: Machine Learning I
## Lecture 29.2: Bias and Variance - Examples

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE UNIVERSITY

©Stanley Chan 2020. All Rights Reserved.

# Outline

- Lecture 28 Sample and Model Complexity
- Lecture 29 Bias and Variance
- Lecture 30 Overfit

**Today's Lecture**:

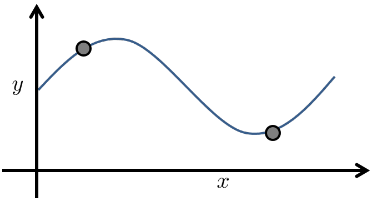
- From VC Analysis to Bias-Variance
  - Generalization Bound
  - Bias-Variance Decomposition
  - Interpreting Bias-Variance
- Example
  - 0-th order vs 1-st order model
  - Trade off

©Stanley Chan 2020. All Rights Reserved.

# Example

- Consider a $\sin(\cdot)$ function

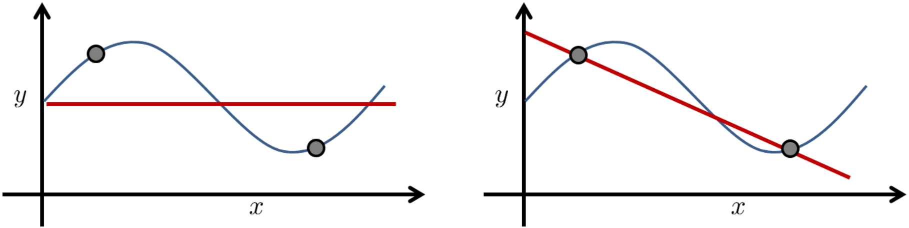
$$f(x) = \sin(\pi x)$$




- You are only given $N = 2$ training samples
- These two samples are sampled uniformly in $[-1, 1]$.
- Call them $(x_1, y_1)$ and $(x_2, y_2)$
- Hypothesis Set 0: $\mathcal{M}_0$ = Set of all lines of the form $h(x) = b$;
- Hypothesis Set 1: $\mathcal{M}_1$ = Set of all lines of the form $h(x) = ax + b$.
- Which one fits better?

©Stanley Chan 2020. All Rights Reserved.

# Example




- If you give me two points, I can tell you the fitted lines
- For $\mathcal{M}_0$:

$$h(x) = \frac{y_1 + y_2}{2}.$$

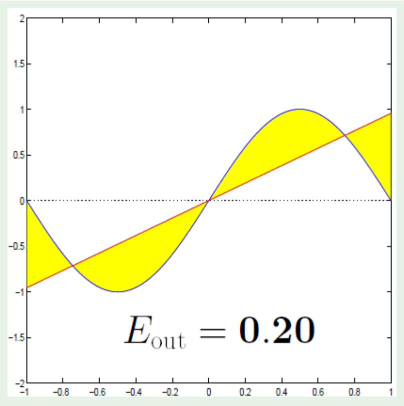
- For $\mathcal{M}_1$:

$$h(x) = \left(\frac{y_2 - y_1}{x_2 - x_1}\right) x + (y_1 x_2 - y_2 x_1).$$

©Stanley Chan 2020. All Rights Reserved.

# Out-sample Error $E_{\text{out}}$

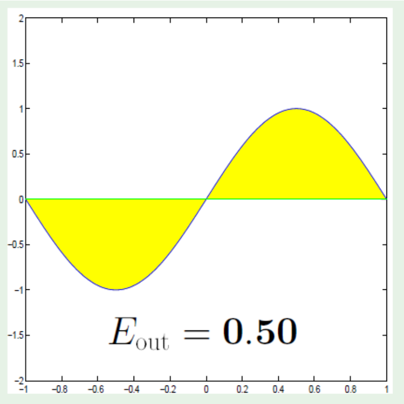
- If you use $\mathcal{M}_1$
- Then you get this
- $E_{\text{out}} = 0.2$




©Stanley Chan 2020. All Rights Reserved.

# Out-sample Error $E_{\text{out}}$

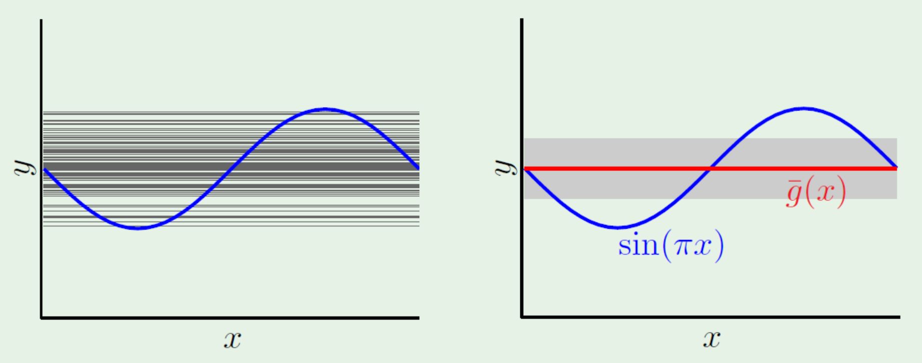
- If you use $\mathcal{M}_0$
- Then you get this
- $E_{\text{out}} = 0.5$




©Stanley Chan 2020. All Rights Reserved.

# Scan through $\mathcal{D}$

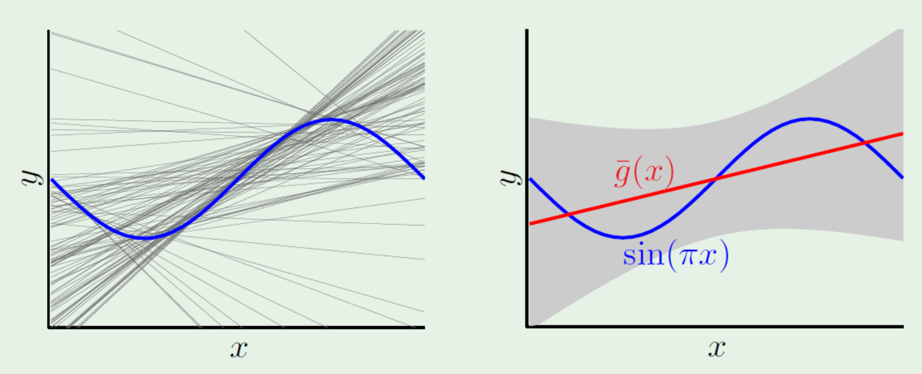
- Now draw a different training set
- Then you have a different curve every time
- Plot them all on the same figure
- Here is what you will get




©Stanley Chan 2020. All Rights Reserved.

# Scan through $\mathcal{D}$

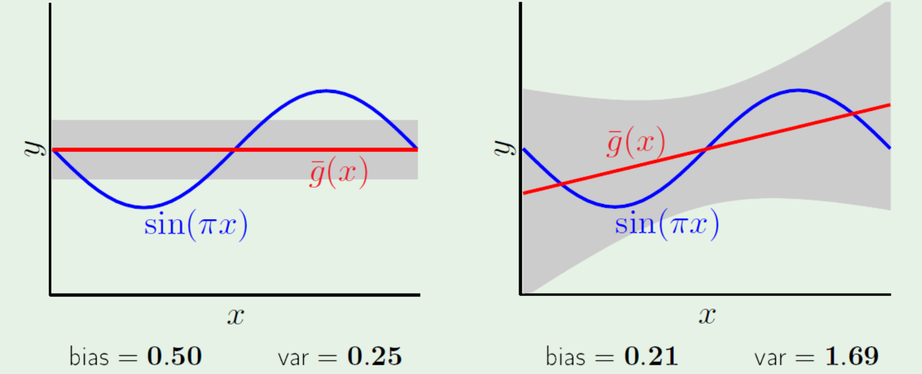
- Now draw a different training set
- Then you have a different curve every time
- Plot them all on the same figure
- Here is what you will get




n 2020. All Rights Reserved.

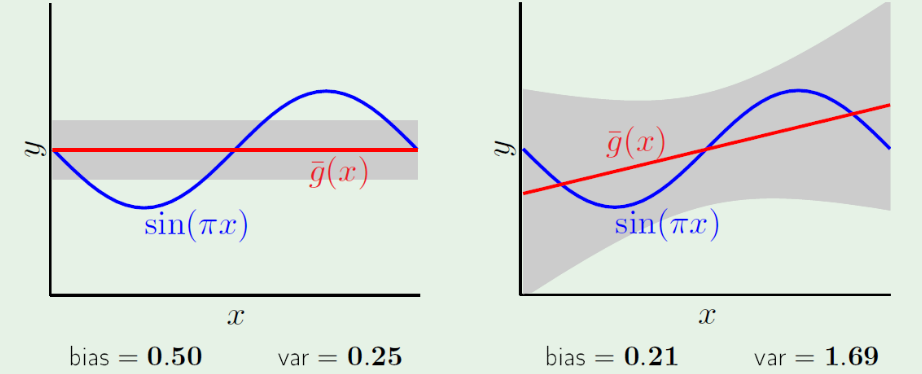
# Limiting Case

- Draw infinitely many training sets
- You will have two quantities
- $\overline{g}(x)$: The average line
- $\sqrt{\text{var}(x)}$: The variance




n 2020. All Rights Reserved.

# How Come!

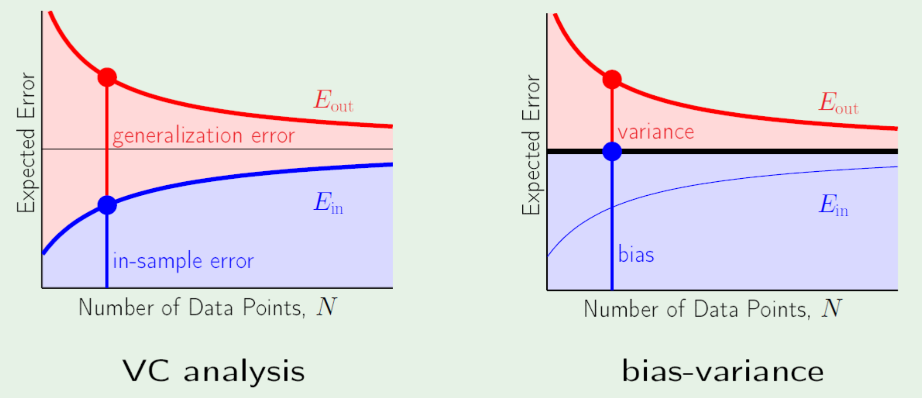



- $\bar{g}(\boldsymbol{x})$ is a good **average**.
- But the **error bar** is big!
- Analogy: I have a powerful canon but not very accurate.

©Stanley Chan 2020. All Rights Reserved.

# Learning Curve

- Expected out-sample error: $E_{\text{out}}(g^{(\mathcal{D})})$
- Expected in-sample error: $E_{\text{in}}(g^{(\mathcal{D})})$
- How do they change with $N$?




©Stanley Chan 2020. All Rights Reserved.

# VC vs Bias-Variance

- VC analysis is independent of $\mathcal{A}$
- Bias-variance depends on $\mathcal{A}$
- With the same $\mathcal{H}$, VC always returns the same generalization bound
- Guarantee over all possible choices of dataset $\mathcal{D}$
- Bias-variance: For the same $\mathcal{H}$, you can have different $g^{(\mathcal{D})}$
- Depend on $\mathcal{D}$, you have a different $E_{\text{out}}(g^{(\mathcal{D})})$
- Therefore we take expectation

$$\mathbb{E}_{\mathcal{D}}\left[E_{\text{out}}(g^{(\mathcal{D})})\right]$$

- In practice, bias and variance cannot be computed
- You do not have $f$
- It is a conceptual tool to design algorithms

© Stanley Chan 2020. All Rights Reserved.

# Reading List

- Yasar Abu-Mostafa, Learning from Data, chapter 2.2
- Chris Bishop, Pattern Recognition and Machine Intelligence, chapter 3.2
- Duda, Hart and Stork, Pattern Classification, chapter 9.3
- Stanford STAT202 https://web.stanford.edu/class/stats202/content/lec2.pdf
- CMU 10-601 https://www.cs.cmu.edu/~wcohen/10-601/bias-variance.pdf
- UCSD 271A http://www.svcl.ucsd.edu/courses/ece271A/handouts/ML2.pdf

©Stanley Chan 2020. All Rights Reserved.