Bias-Variance Tradeoff

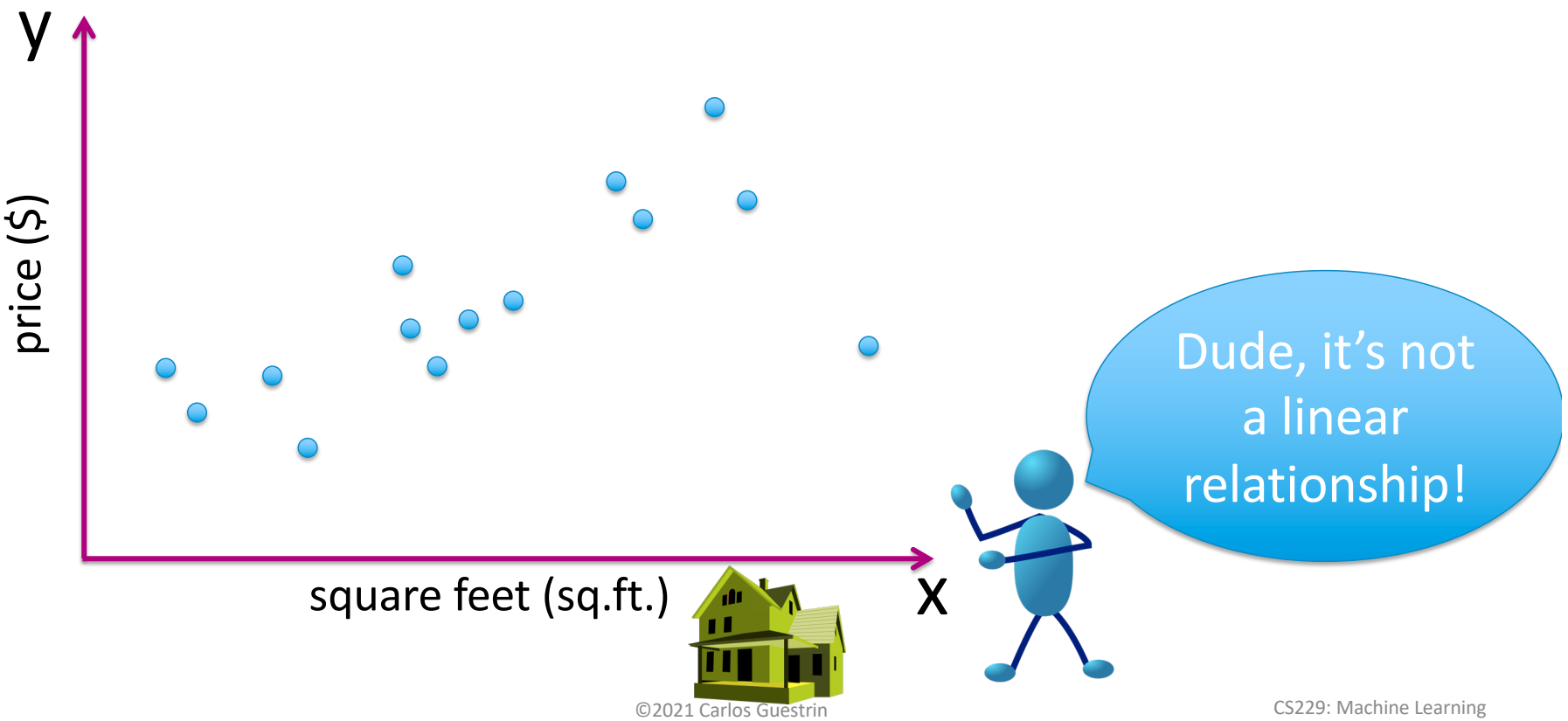
CS229: Machine Learning

Carlos Guestrin
Stanford University

Slides include content developed by and co-developed with Emily Fox

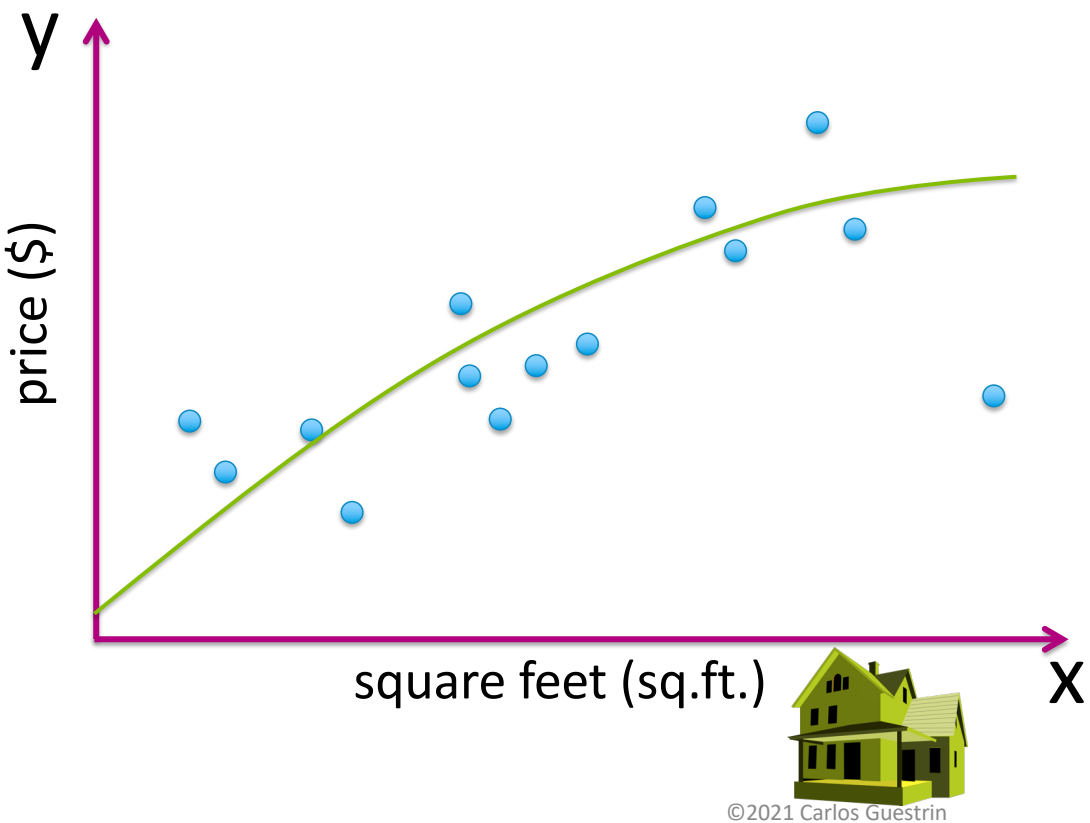
©2021 Carlos Guestrin

Fit data with a line or … ?



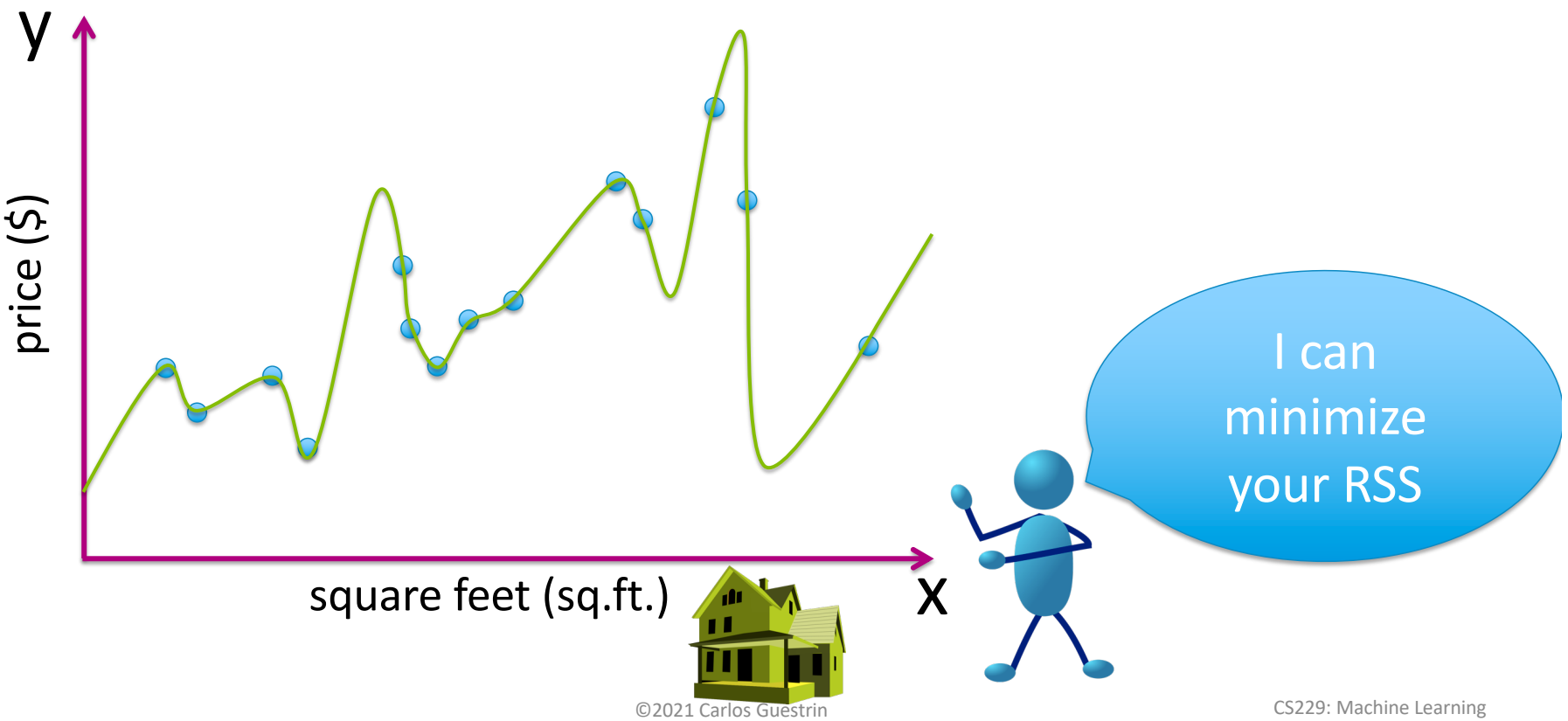
©2021 Carlos Guestrin

What about a quadratic function?



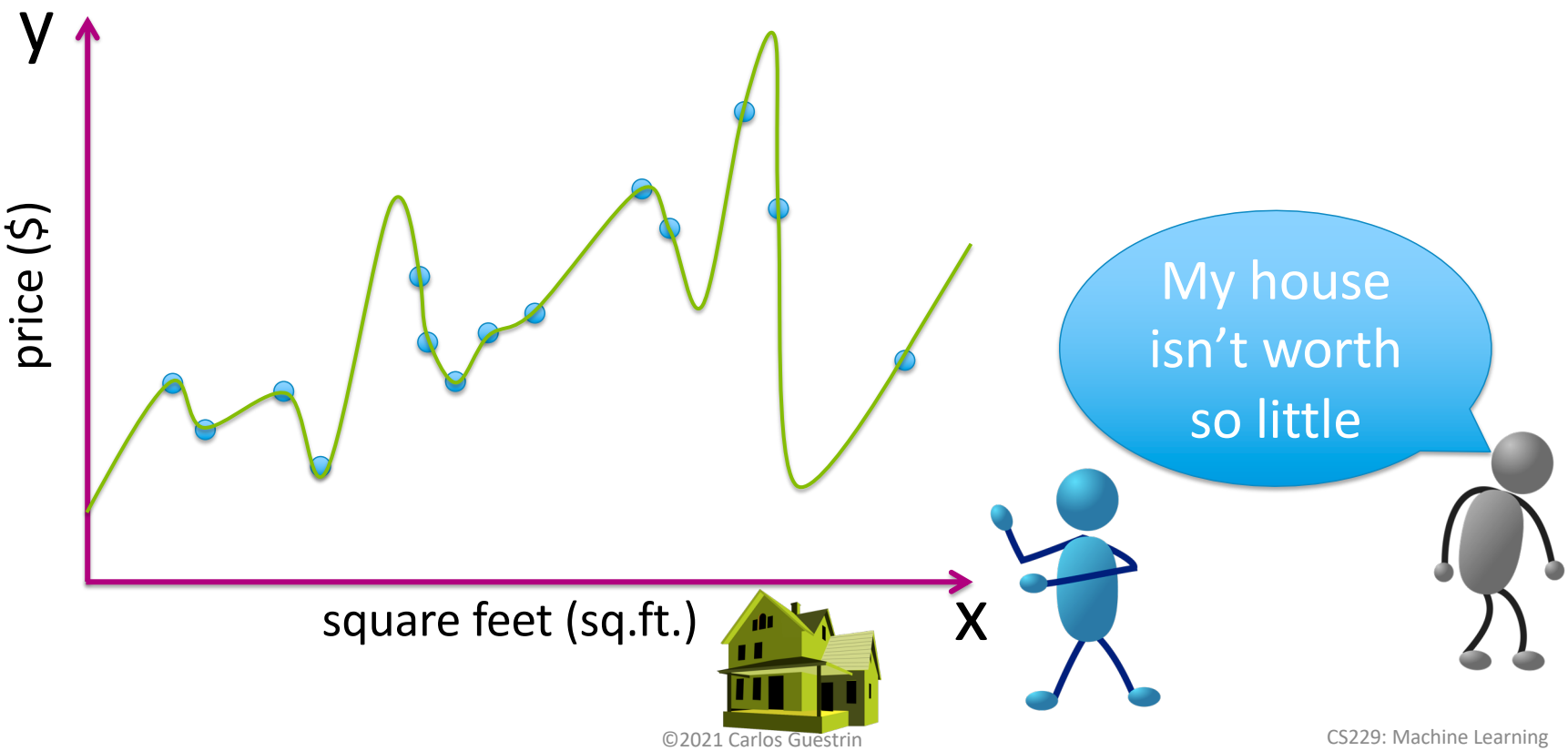
©2021 Carlos Guestrin

Even higher order polynomial



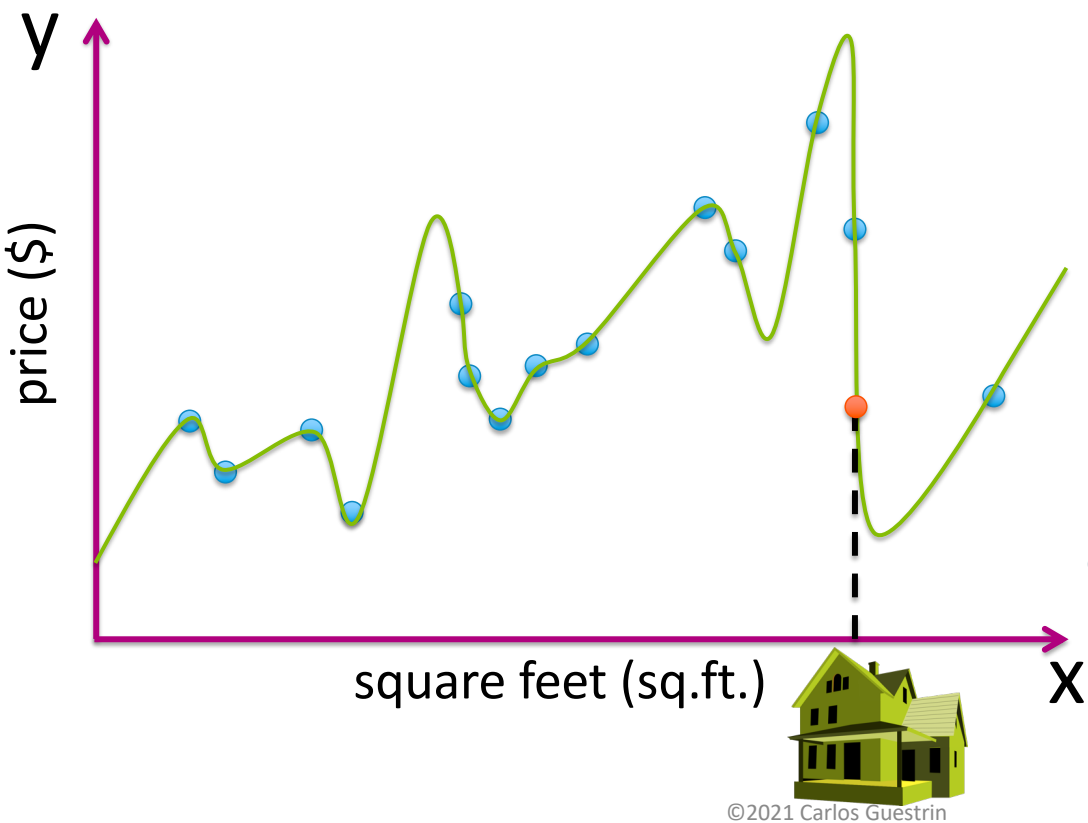
©2021 Carlos Guestrin

Do you believe this fit?

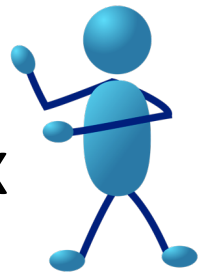


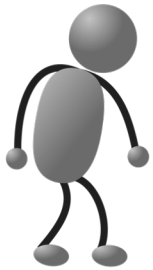
©2021 Carlos Guestrin

Do you believe this fit?



Minimizes RSS,
but bad predictions





©2021 Carlos Guestrin

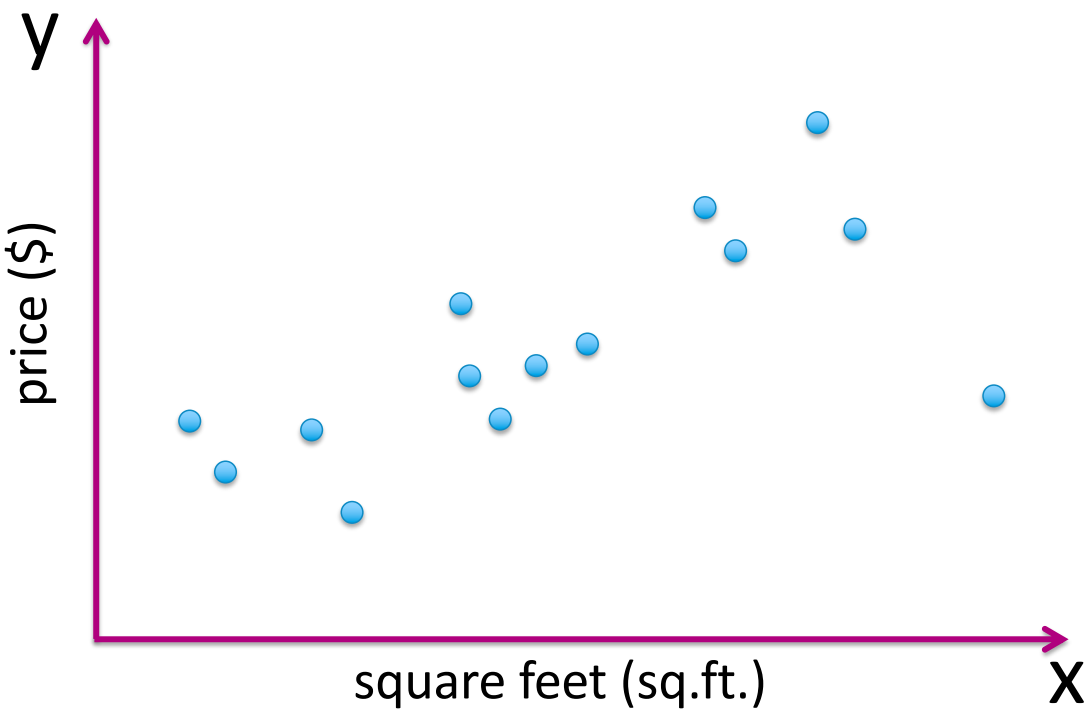
"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." George Box, 1987.

Assessing the loss

Assessing the loss
Part 1: Training error

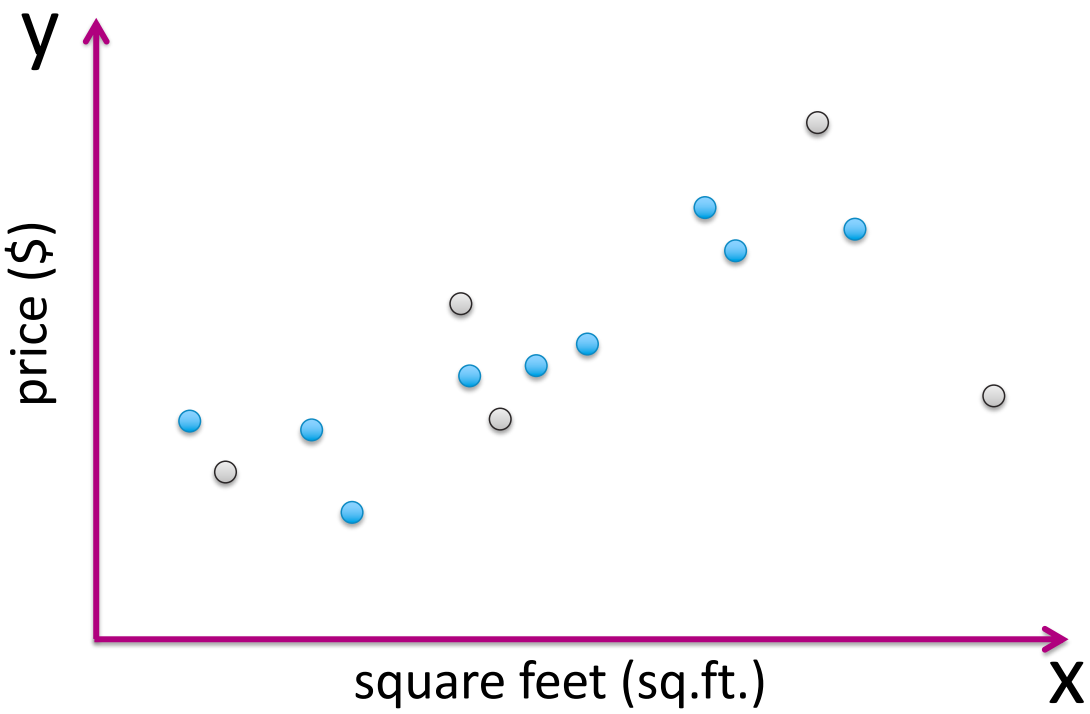
©2021 Carlos Guestrin

Define training data



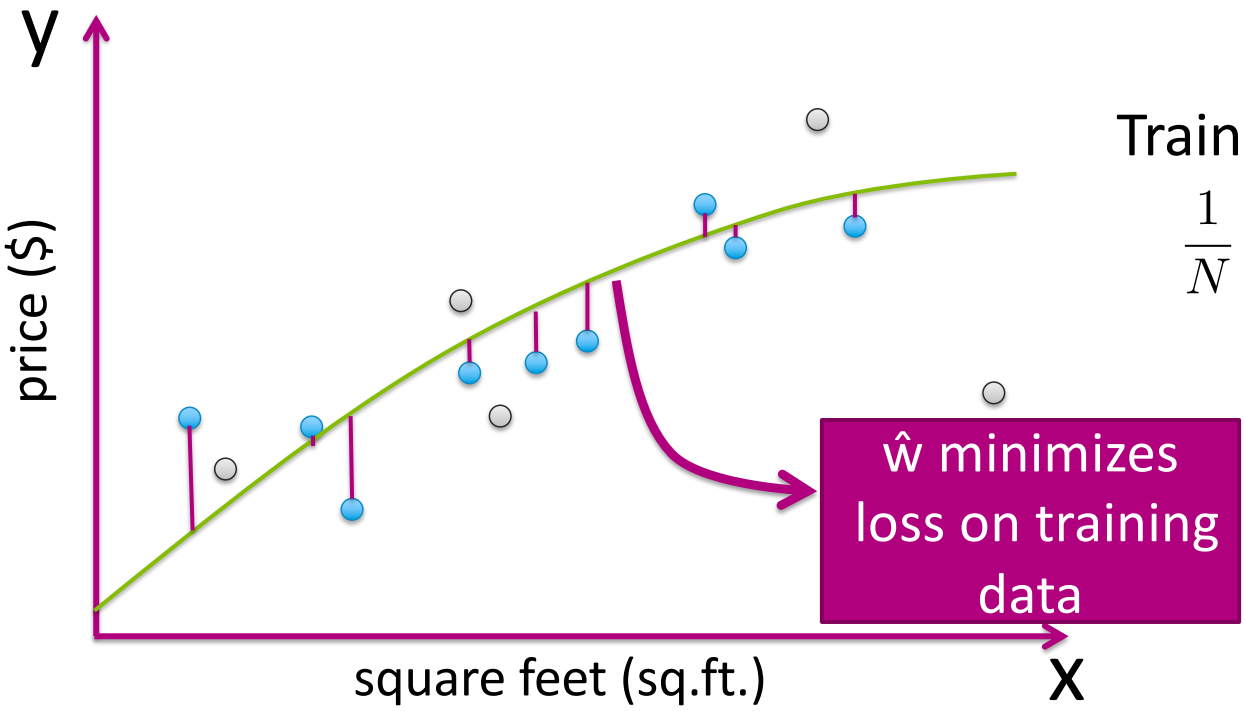
©2021 Carlos Guestrin

Define training data



©2021 Carlos Guestrin

Example:
Fit quadratic to minimize RSS



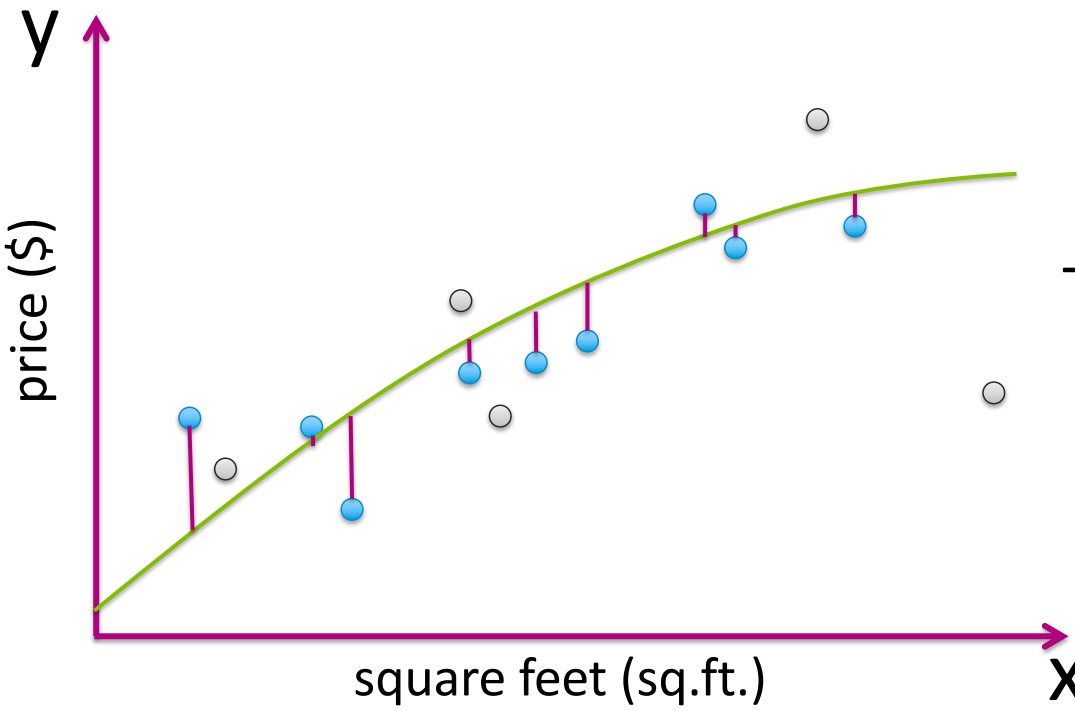
Training error ($\hat{w}$) =

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - f_{\hat{w}}(x_i))^2$$

©2021 Carlos Guestrin

Example:

Use squared error loss $(y-f_{\hat{w}}(x))^2$



Training error ($\hat{w}$) = 1/N *

$[(\$_{\text{train 1}}-f_{\hat{w}}(\text{sq.ft.}_{\text{train 1}}))^2$

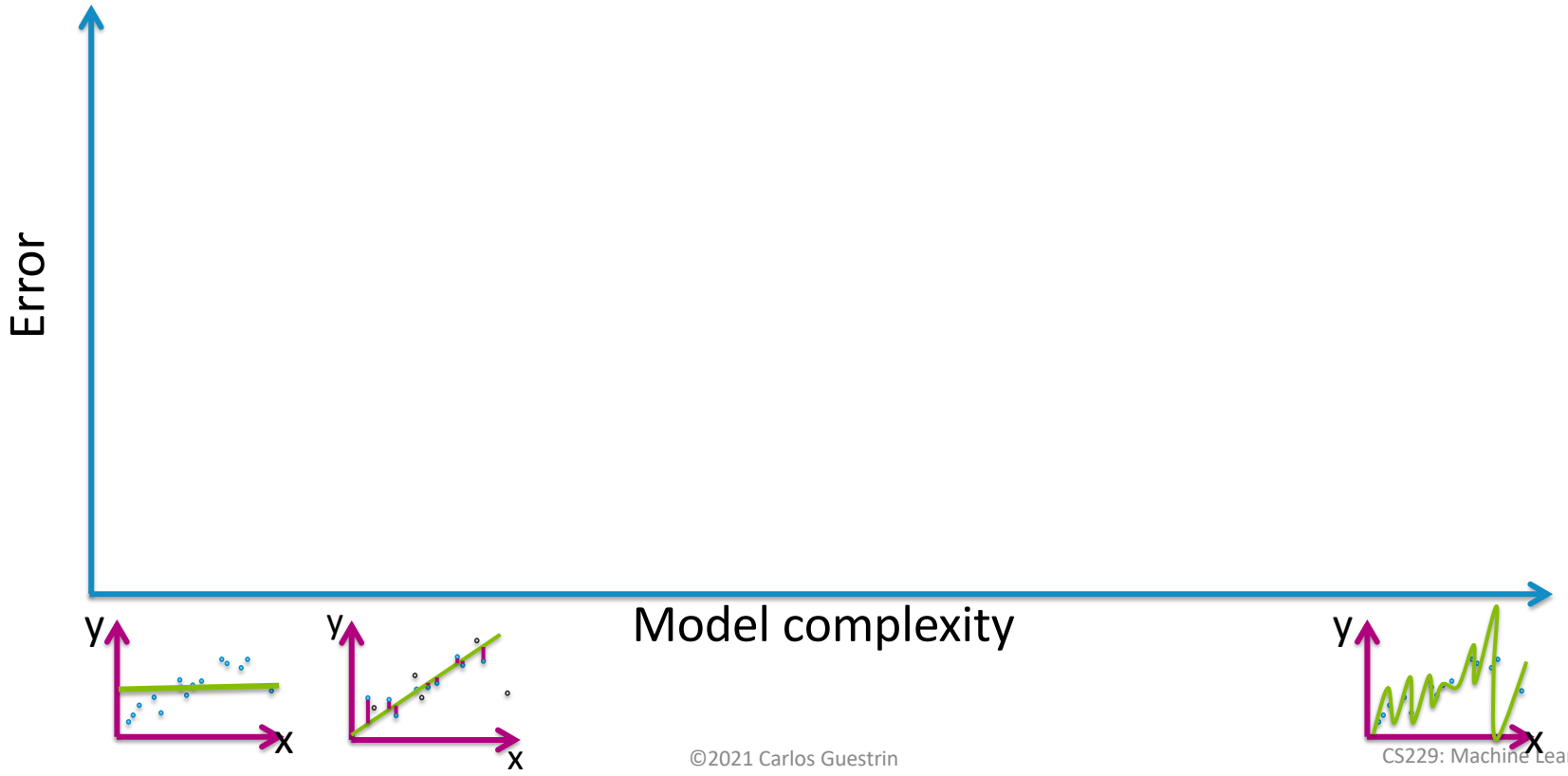
$+ (\$_{\text{train 2}}-f_{\hat{w}}(\text{sq.ft.}_{\text{train 2}}))^2$

$+ (\$_{\text{train 3}}-f_{\hat{w}}(\text{sq.ft.}_{\text{train 3}}))^2$

+ … include all training houses]

©2021 Carlos Guestrin

Training error vs. model complexity

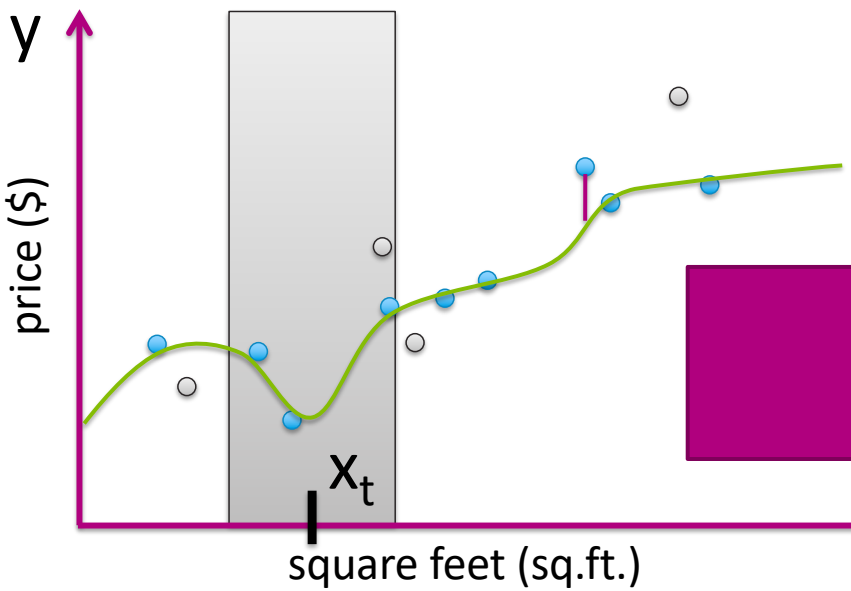


©2021 Carlos Guestrin

Is training error a good measure of predictive performance?

Issue:

Training error is overly optimistic…$\hat{w}$ was fit to training data



Small training error ≠ good predictions
(unless training data includes
everything you might ever see)

©2021 Carlos Guestrin

Assessing the loss
Part 2: Generalization (true) error

©2021 Carlos Guestrin

Generalization error

Really want estimate of loss over all possible (🏠,$) pairs

Lots of houses
in neighborhood,
but not in dataset

©2021 Carlos Guestrin

Generalization error definition

Really want estimate of loss over all possible (,$) pairs

Formally:

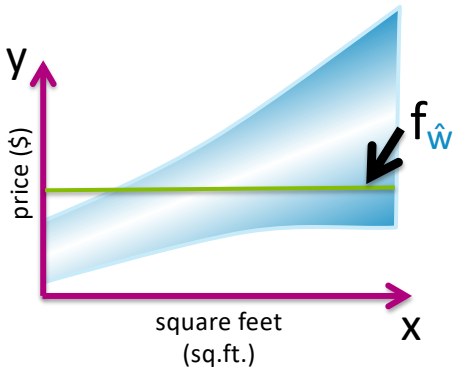
average over all possible (x,y) pairs weighted by how likely each is

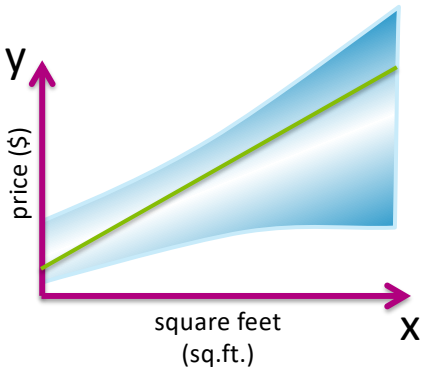
$$\text{generalization error} = E_{x,y}\left[L(y, f_{\hat{w}}(x))\right]$$

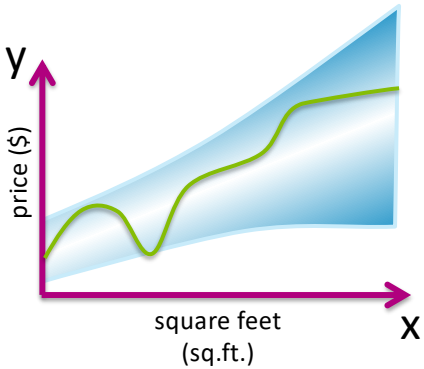
fit using training data

©2021 Carlos Guestrin

Generalization error vs. model complexity

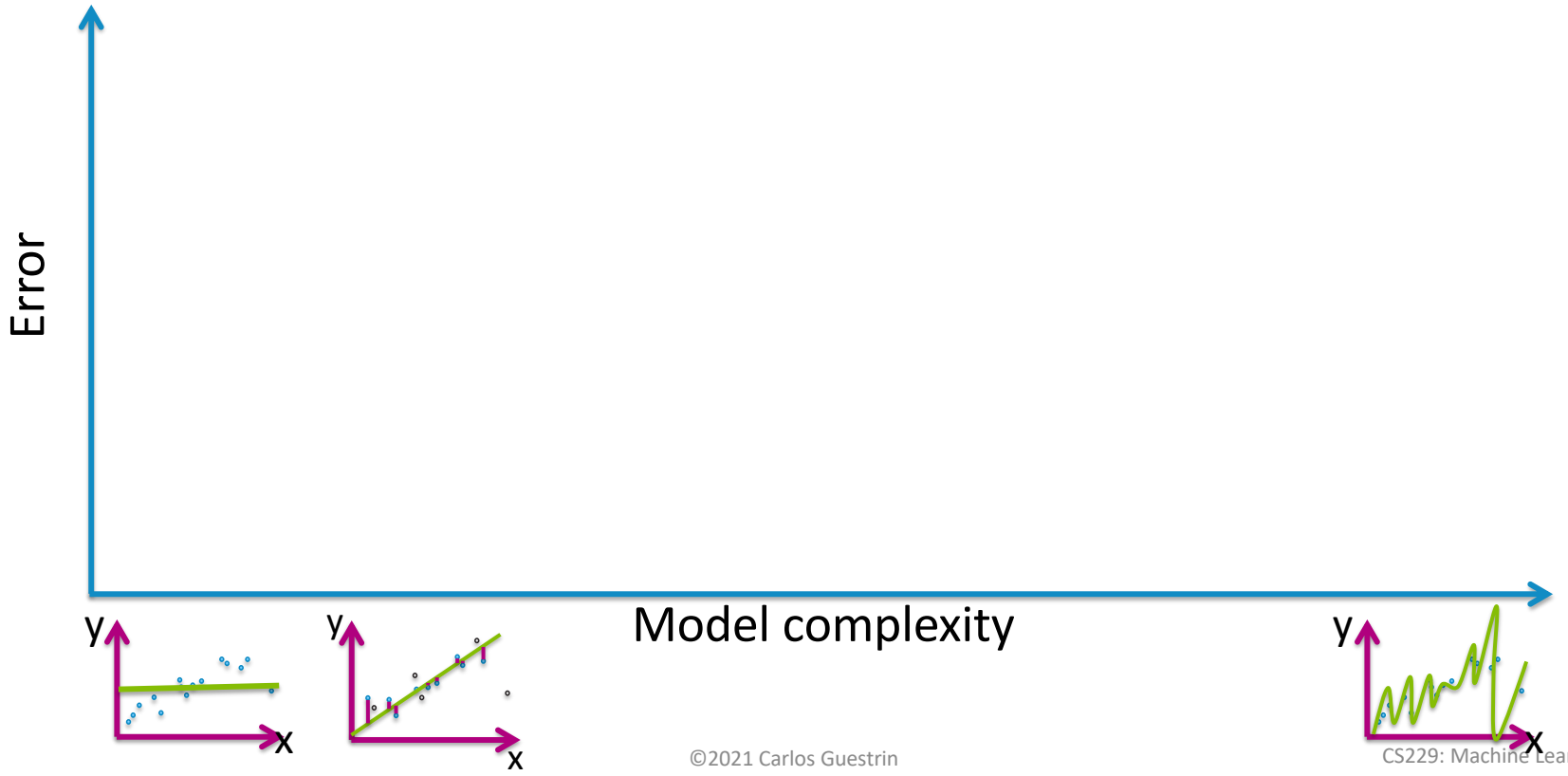






©2021 Carlos Guestrin

True error vs. model complexity



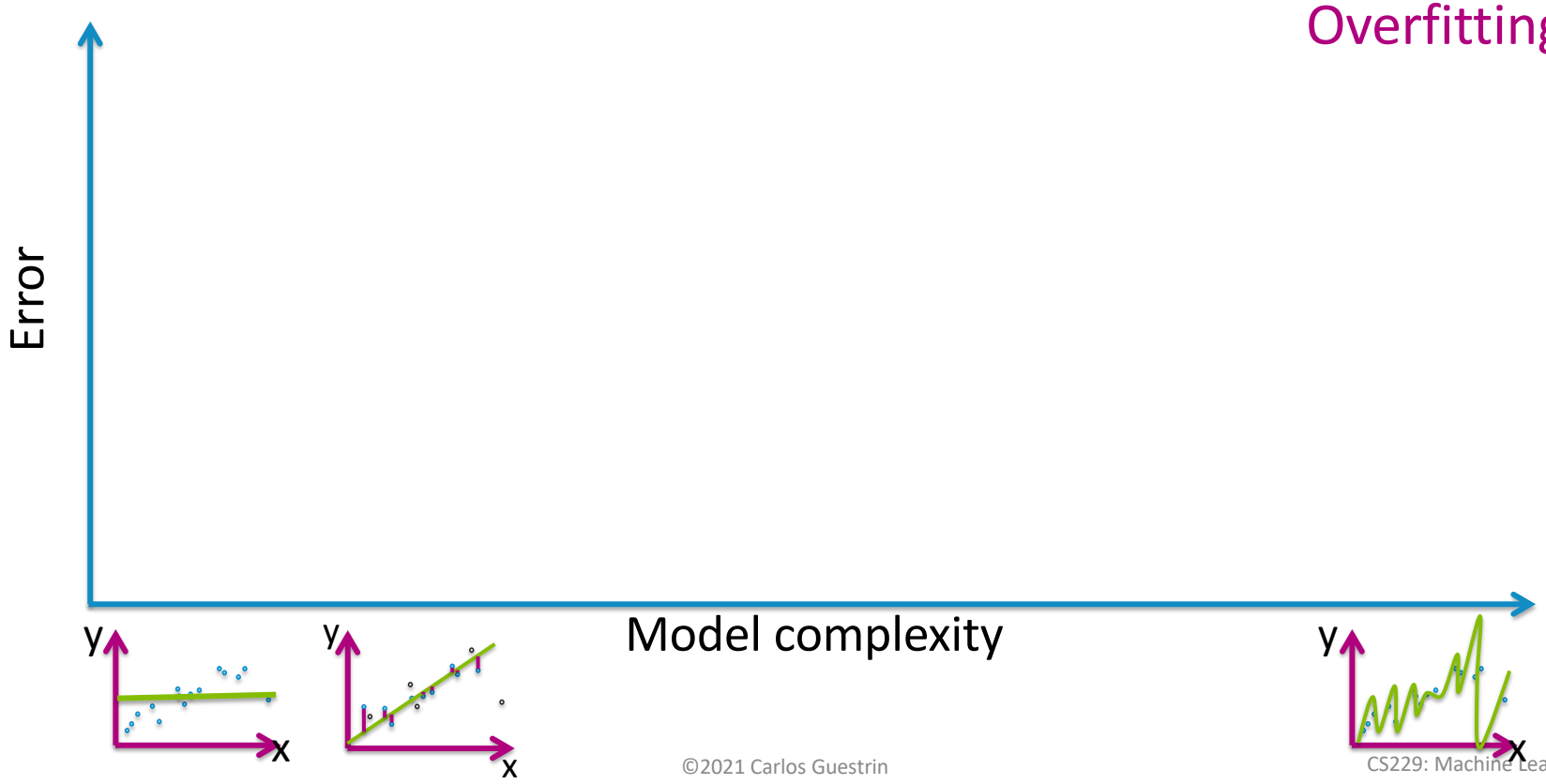
©2021 Carlos Guestrin

Assessing the loss
Part 3: Test error

©2021 Carlos Guestrin

Training, true, test error vs. model complexity

Overfitting if:



©2021 Carlos Guestrin

3 sources of error + the bias-variance tradeoff

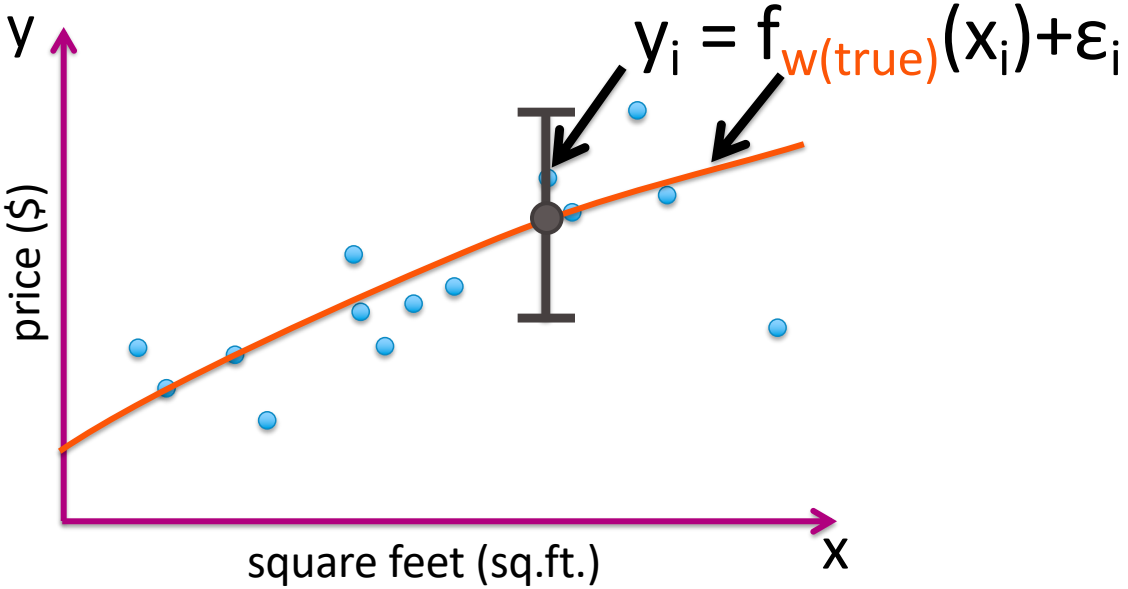
3 sources of error

In forming predictions, there are 3 sources of error:

1. Noise
2. Bias
3. Variance

©2021 Carlos Guestrin

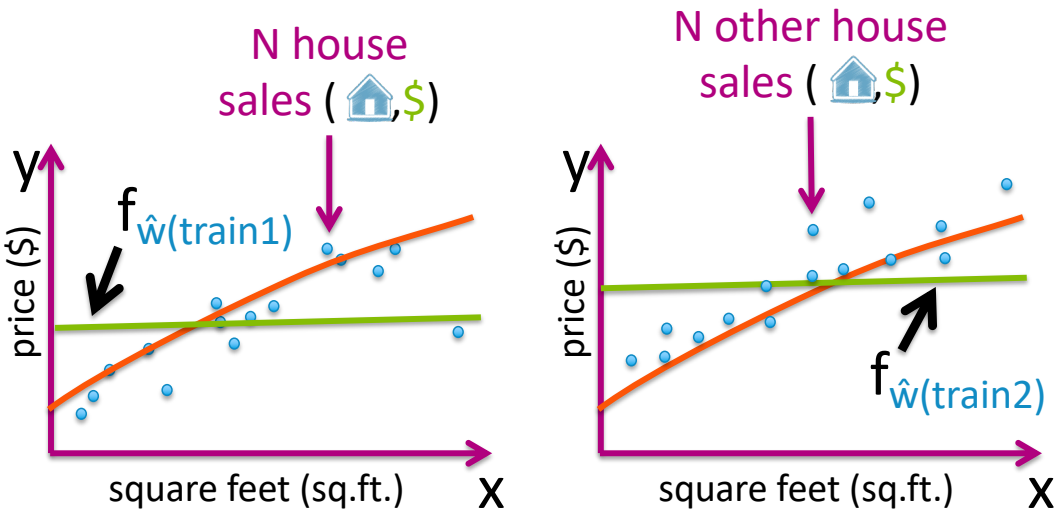
Data inherently noisy



©2021 Carlos Guestrin

Bias contribution

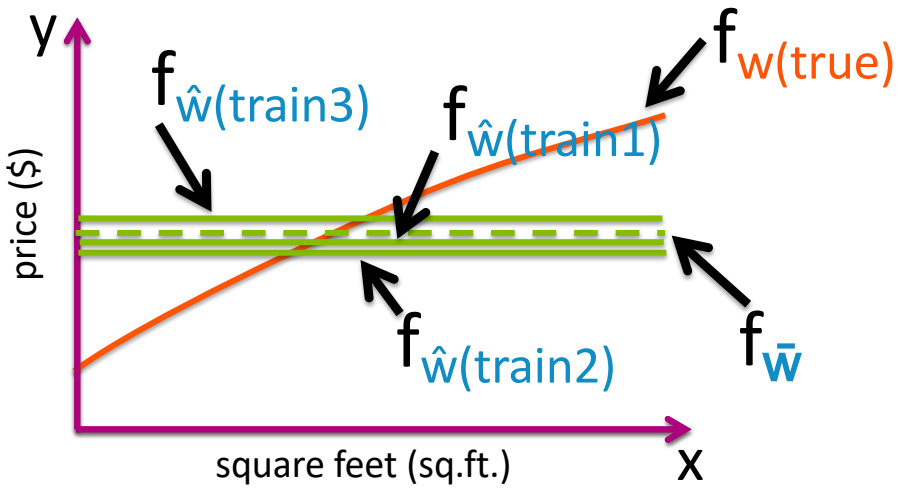
Suppose we fit a constant function



©2021 Carlos Guestrin

Bias contribution

Over all possible size N training sets,
what do I expect my fit to be?

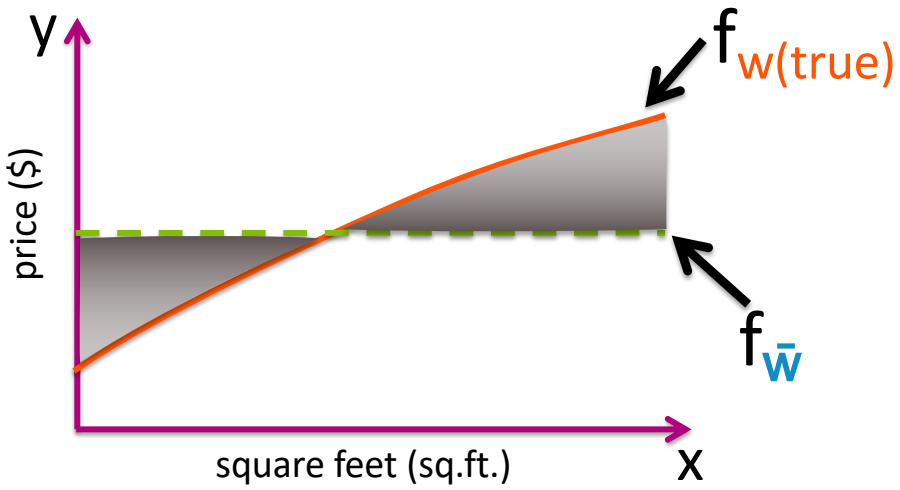


©2021 Carlos Guestrin

Bias contribution

$\text{Bias}(x) = f_{w(\text{true})}(x) - f_{\bar{w}}(x)$

← Is our approach flexible enough to capture $f_{w(\text{true})}$?
If not, error in predictions.

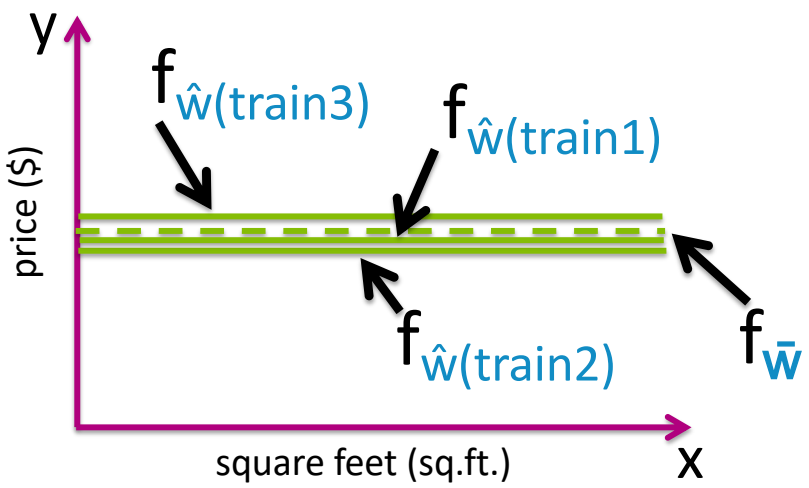


low complexity →

©2021 Carlos Guestrin

Variance contribution

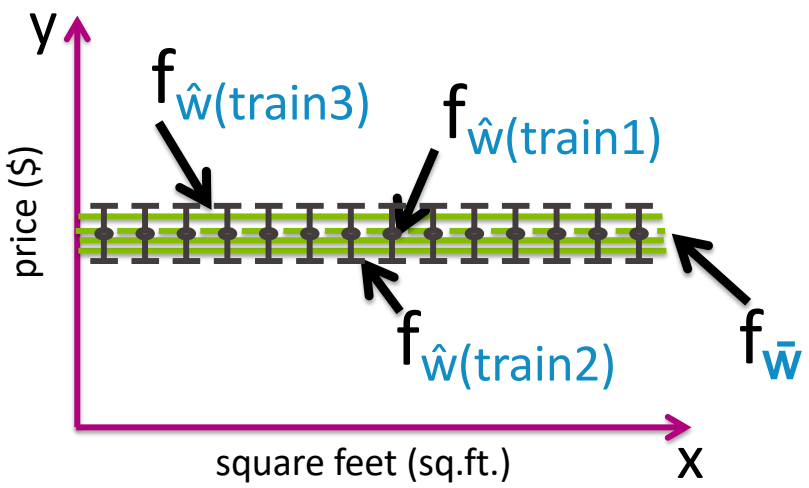
How much do specific fits vary from the expected fit?



©2021 Carlos Guestrin

Variance contribution

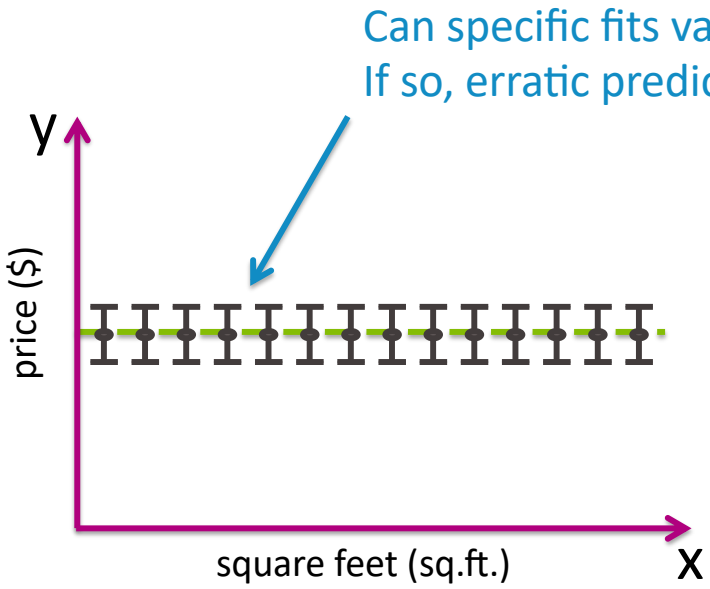
How much do specific fits vary from the expected fit?



©2021 Carlos Guestrin

Variance contribution

How much do specific fits vary from the expected fit?

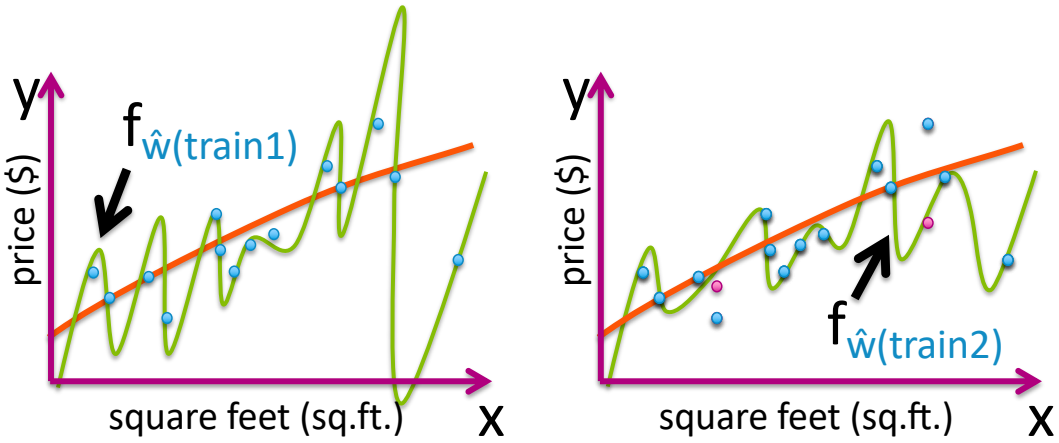


low complexity
→

©2021 Carlos Guestrin

Variance of high-complexity models

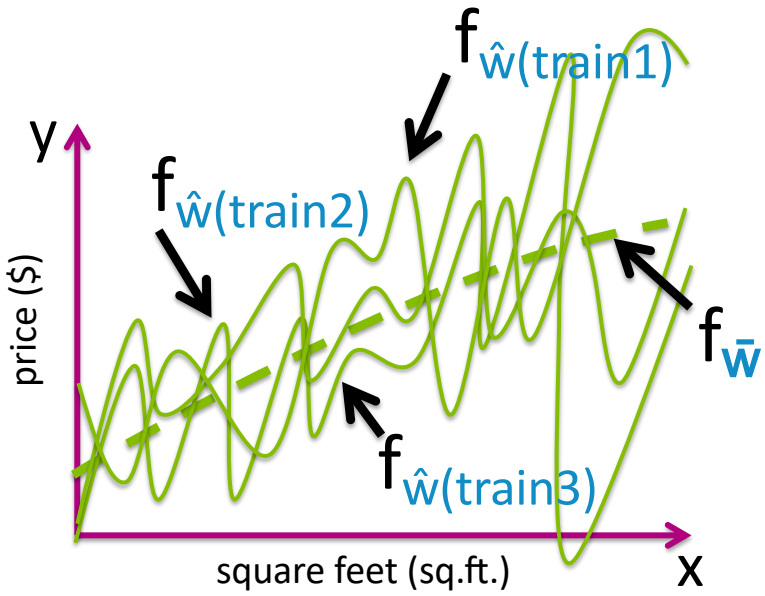
Assume we fit a high-order polynomial



©2021 Carlos Guestrin

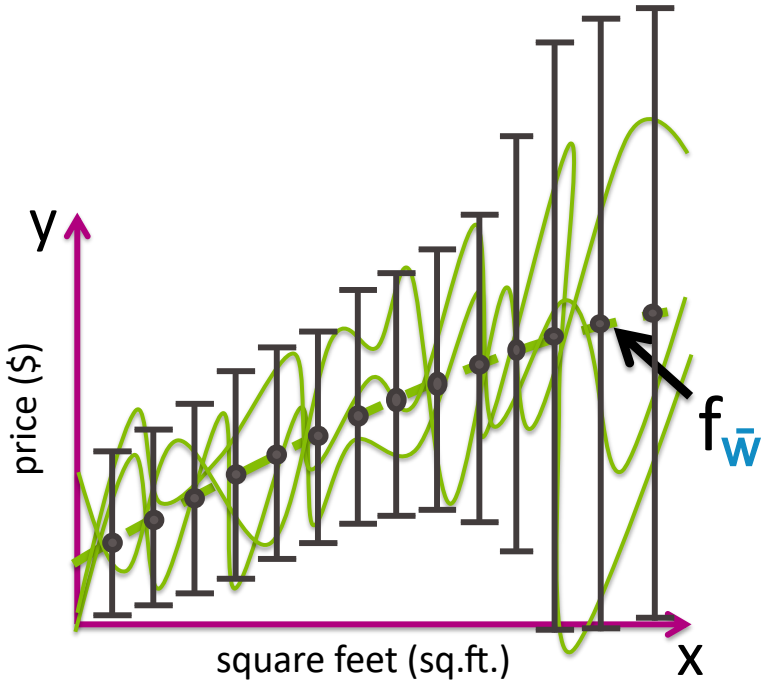
Variance of high-complexity models

Suppose we fit a high-order polynomial



©2021 Carlos Guestrin

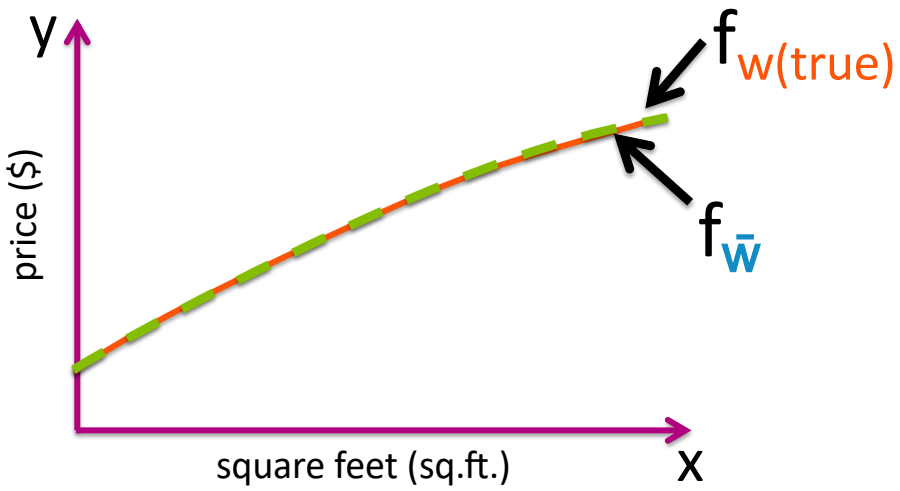
Variance of high-complexity models



high complexity
→

©2021 Carlos Guestrin

Bias of high-complexity models



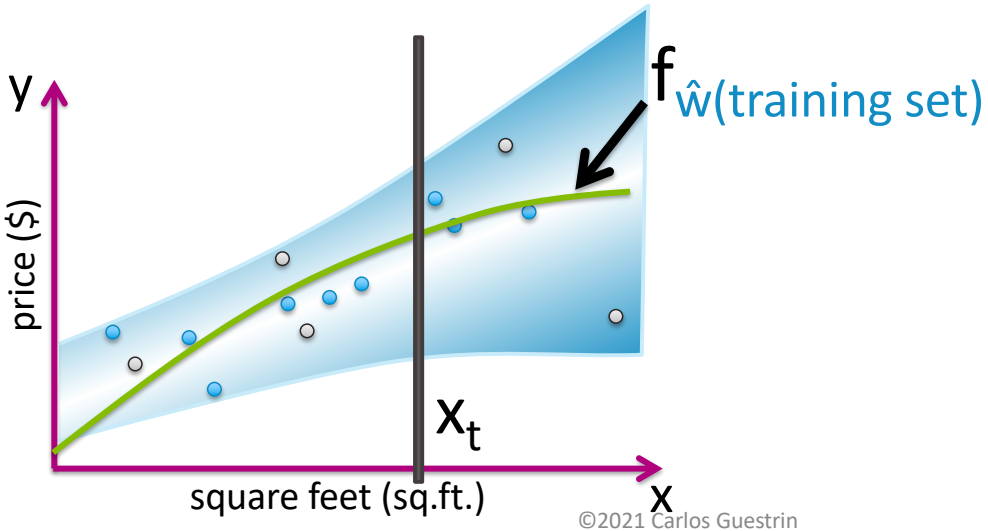
high complexity
→

©2021 Carlos Guestrin

Sum of 3 sources of error

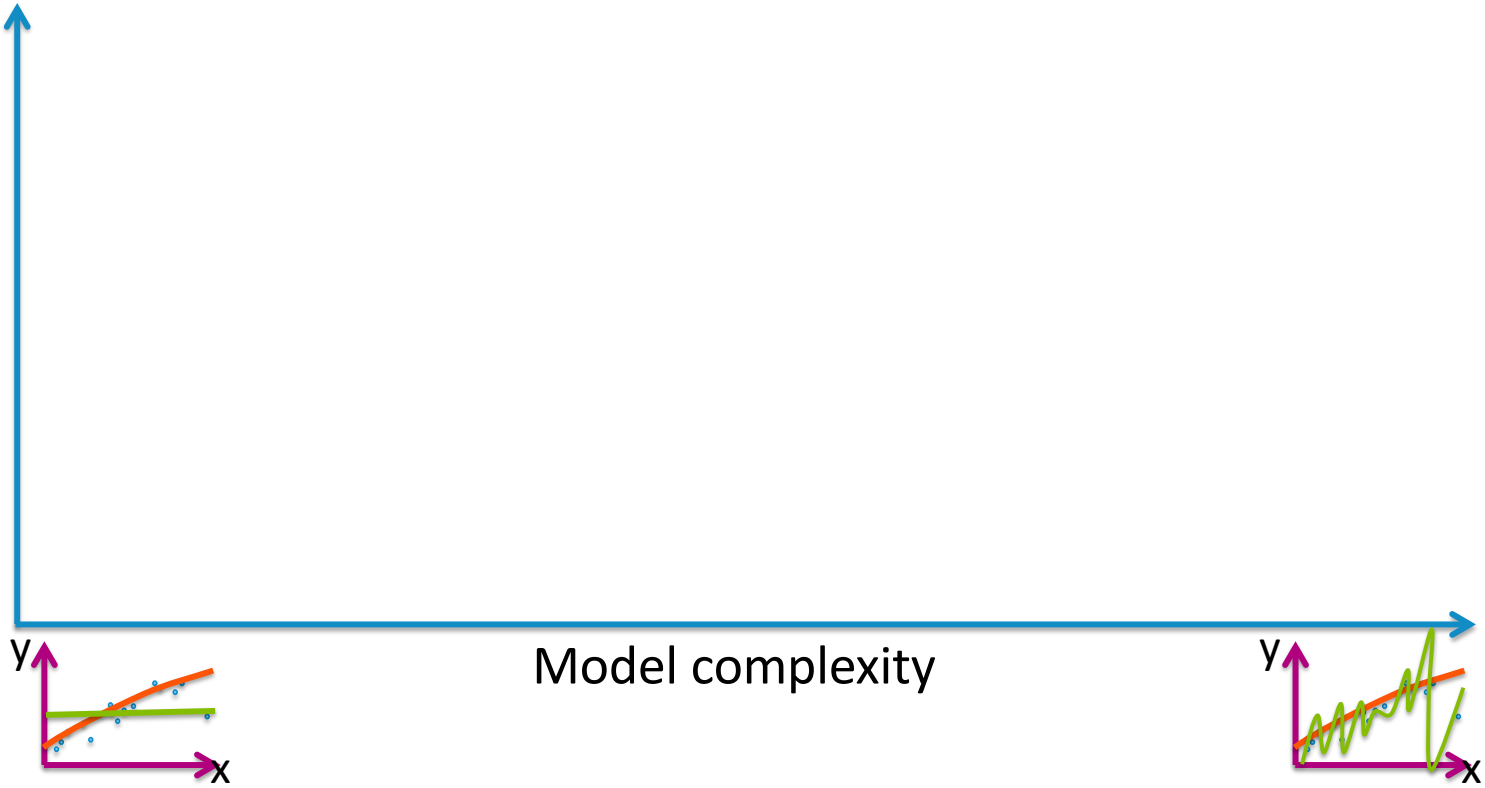
Average squared error at x_t

$$= \sigma^2 + [\text{bias}(f_{\hat{w}}(x_t))]^2 + \text{var}(f_{\hat{w}}(x_t))$$



©2021 Carlos Guestrin

Bias-variance tradeoff



©2021 Carlos Guestrin

Error vs. amount of data

©2021 Carlos Guestrin

Summary of bias-variance tradeoff

What you can do now…

- Contrast relationship between model complexity and train, true and test loss
- Compute training and test error given a loss function for different model complexities
- List and interpret the 3 sources of avg. prediction error
  - Irreducible error, bias, and variance

©2021 Carlos Guestrin