MLDM

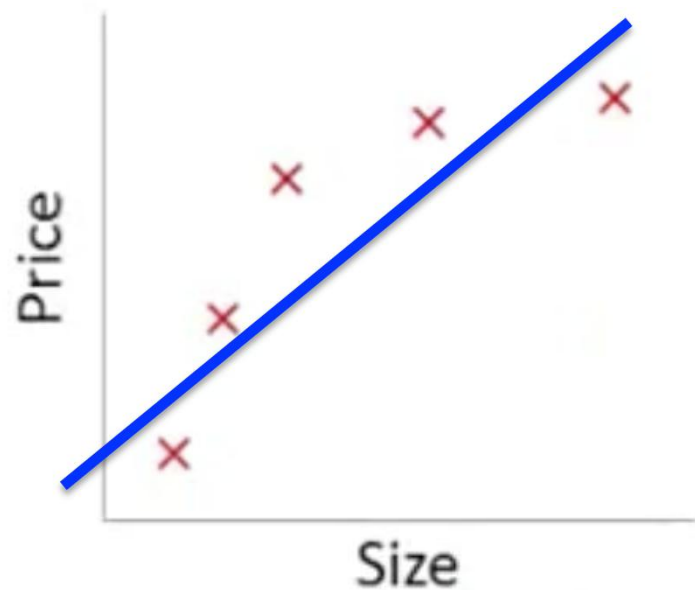# The problem of overfitting

Shangsong Liang

Sun Yat-sen University

- What is overfitting

- How to avoid overfitting
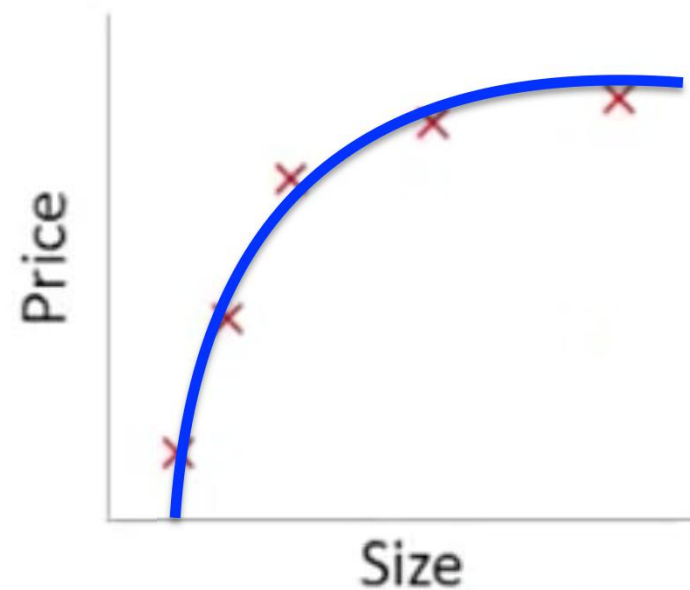
✖ What is overfitting

✖ How to avoid overfitting

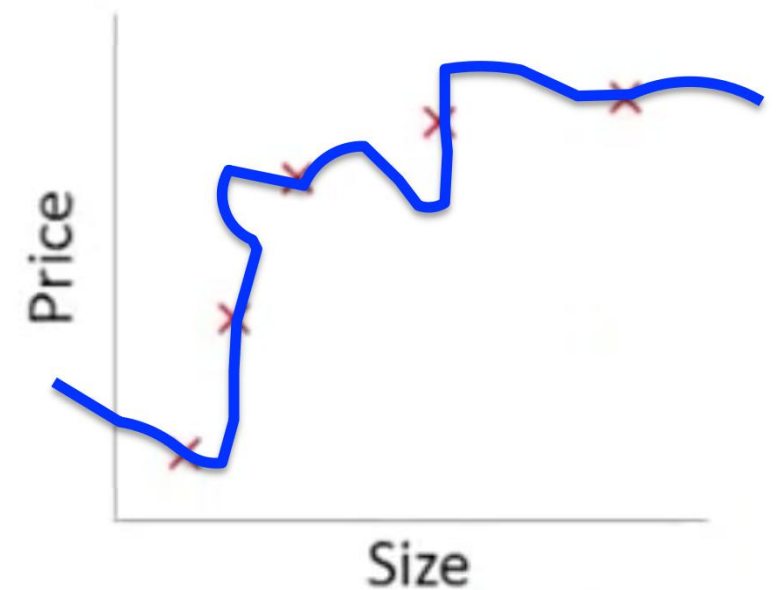# Example: Linear regression (housing prices)

$$\theta_0 + \theta_1 x$$

Underfitting

$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Just right

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

Overfitting

Example: Linear regression (housing prices)

$$E(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

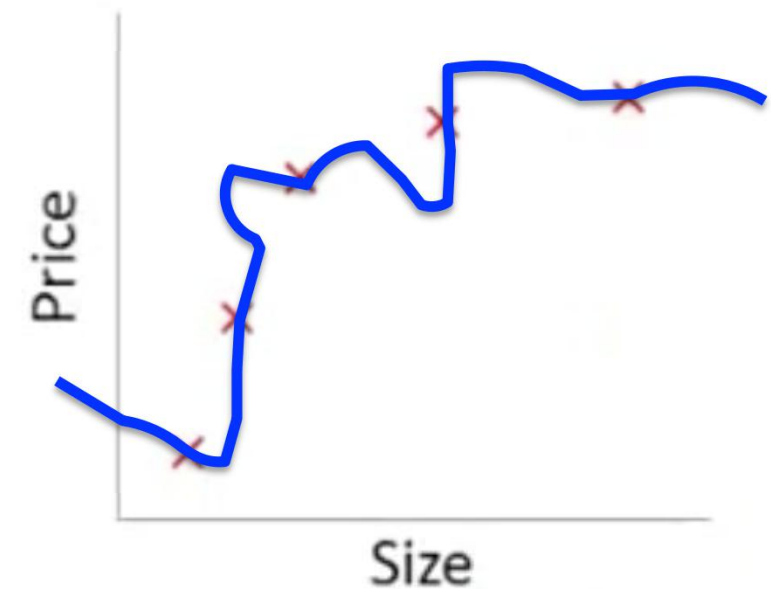|		Observed	Unobserved
$\theta_0 + \theta_1 x$	Underfitting	$E(\theta)$ ↑	↑
$\theta_0 + \theta_1 x + \theta_2 x^2$	Just right	$E(\theta)$ ↓	↓
$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$	Overfitting	$E(\theta)$ ↓	↑

Overfitting: fit the training set very well, but fail to generalize to new examples.

✖ What is overfitting

✖ **How to avoid overfitting**

Addressing overfitting

$x_1 = $ size of house
$x_2 = $ no. of bedrooms
$x_3 = $ no. of floors
$x_4 = $ age of house
$x_5 = $ average income in neighborhood
$x_6 = $ kitchen size

$\vdots$

x_{100}

Addressing overfitting

- ✖ Reduce number of features

- ✖ Keep all the features, but reduce values of parameters

✖ The problem of overfitting

✖ Shangsong Liang

✖ Sun Yat-sen University