

# Learning Curves

*Bias and Variance*

Advice for Applying Machine Learning

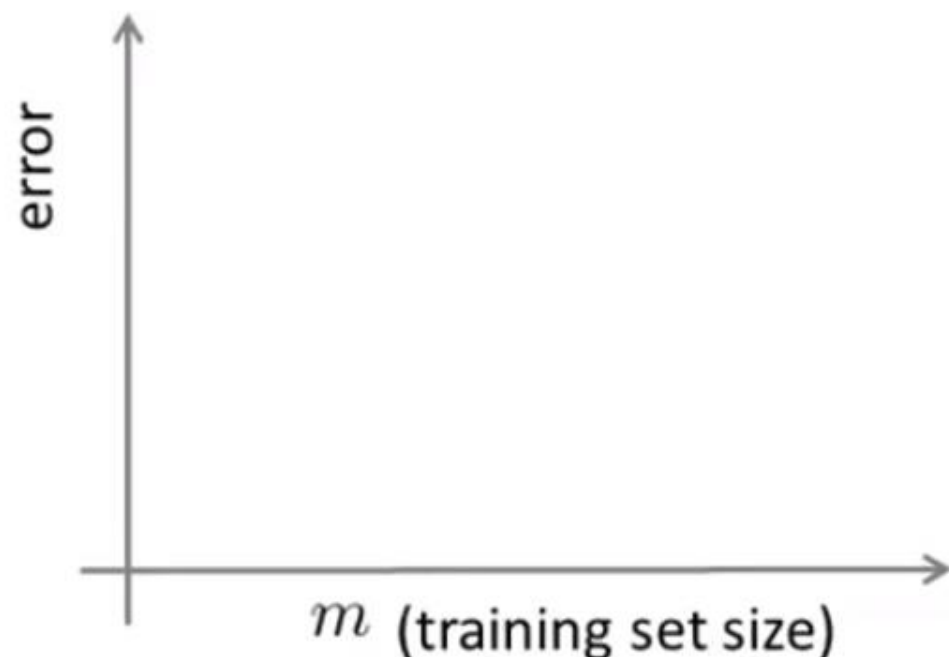
# Introduction

- Learning curves is often a very useful thing to plot.
  - you wanted to sanity check that your algorithm is working correctly, or
  - if you want to improve the performance of the algorithm.
- Tool to diagnose if a physical learning algorithm may be suffering from
  - bias,
  - variance problem or
  - a bit of both.

## Learning curves

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

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

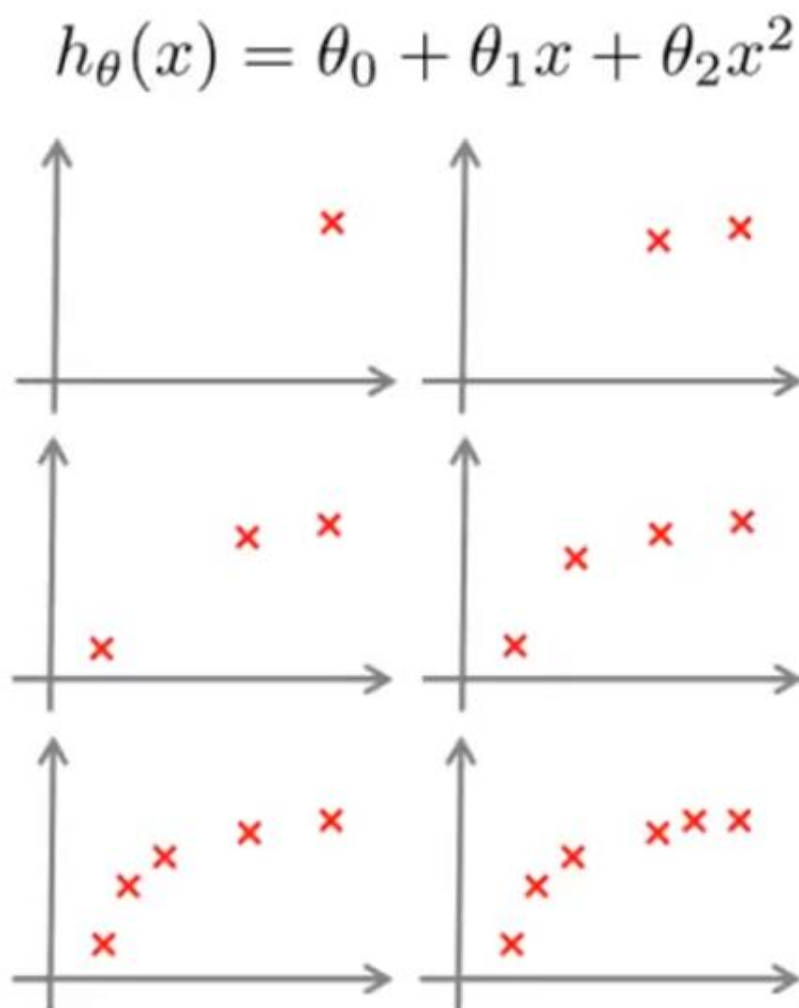
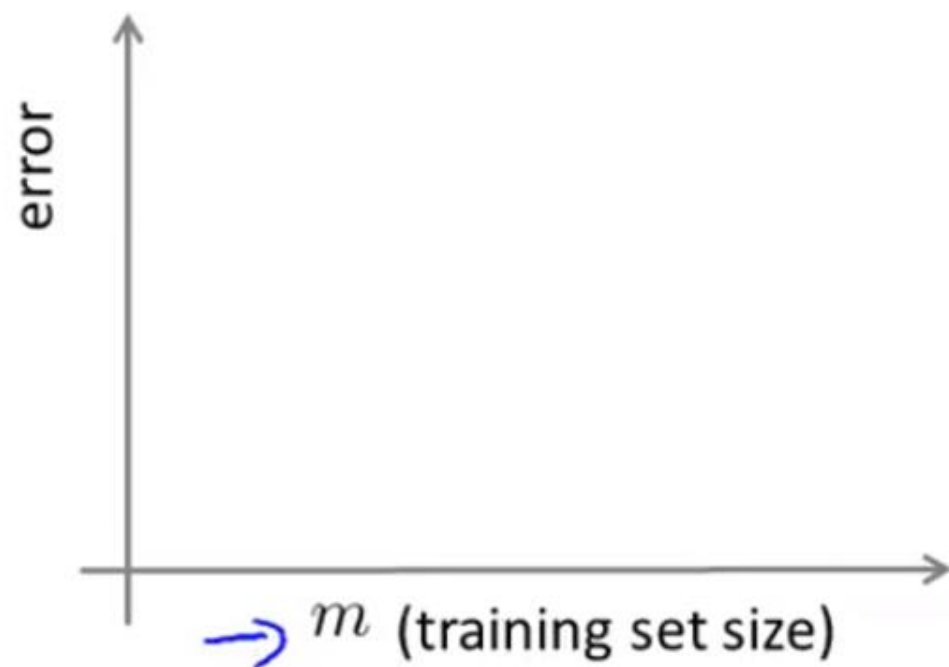


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

## Learning curves

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

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

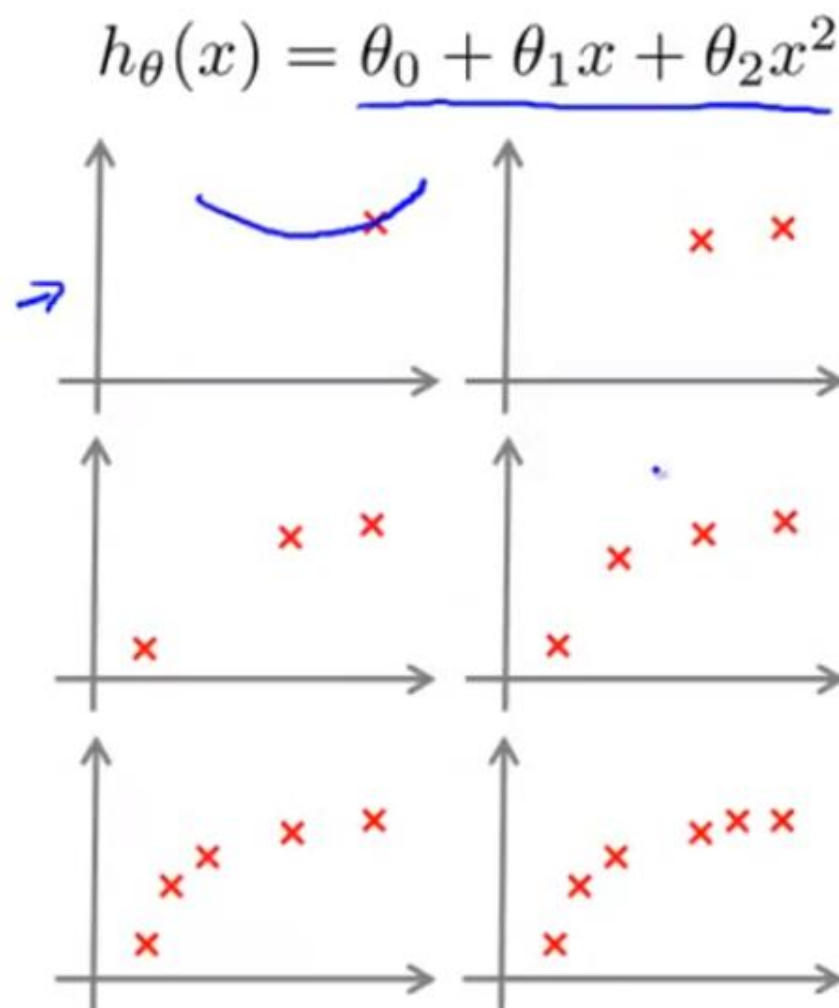
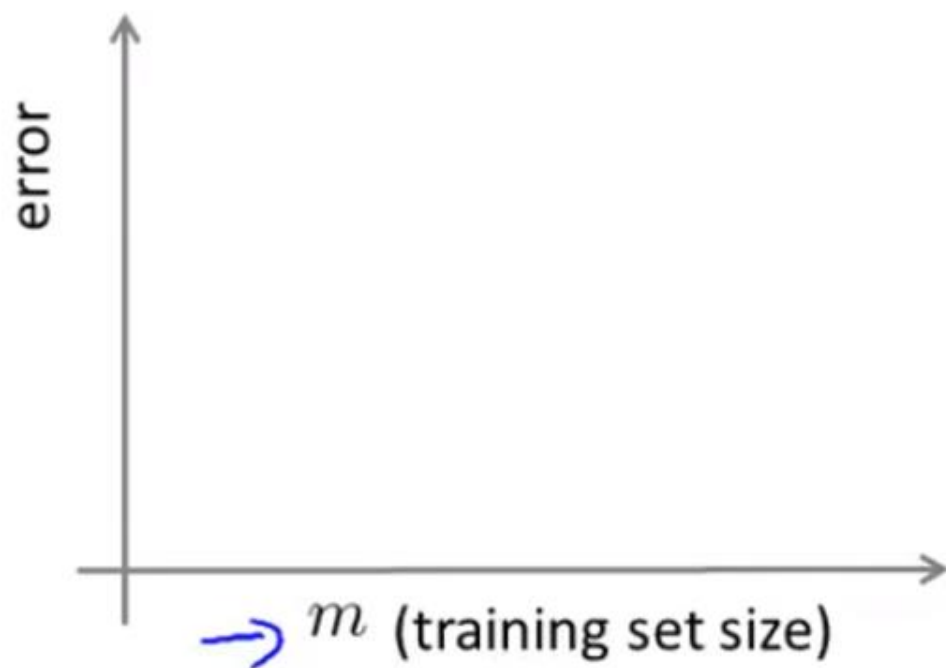


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

# Learning curves

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

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

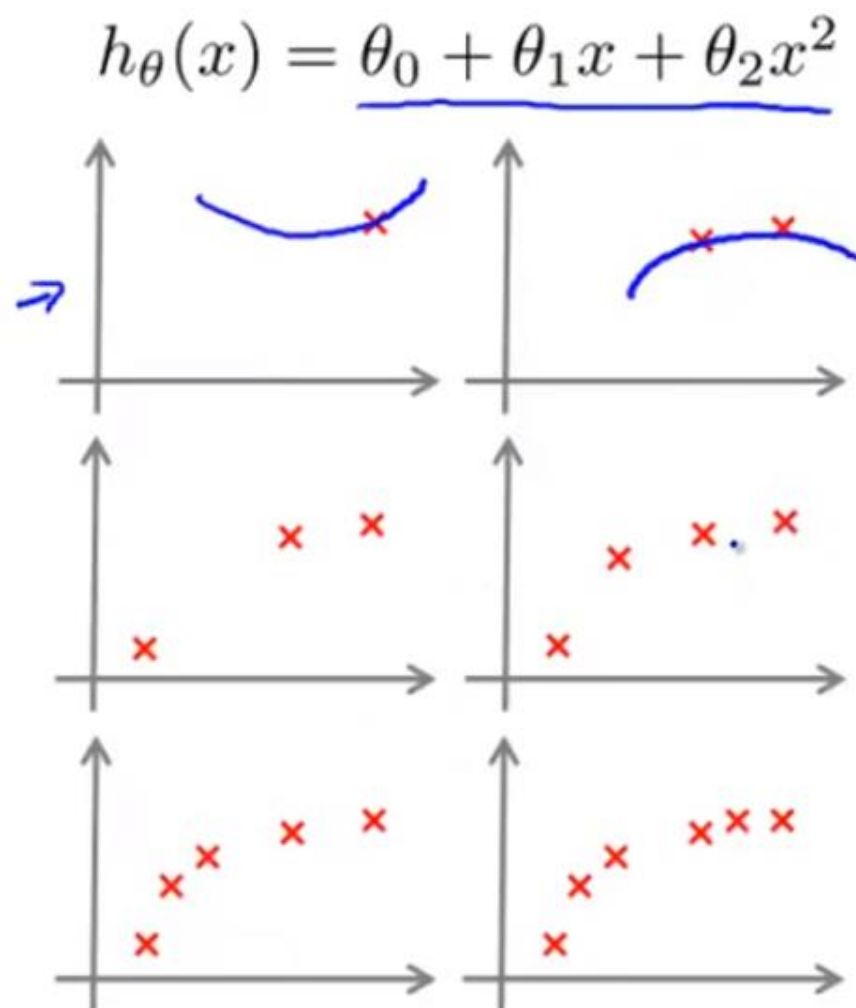
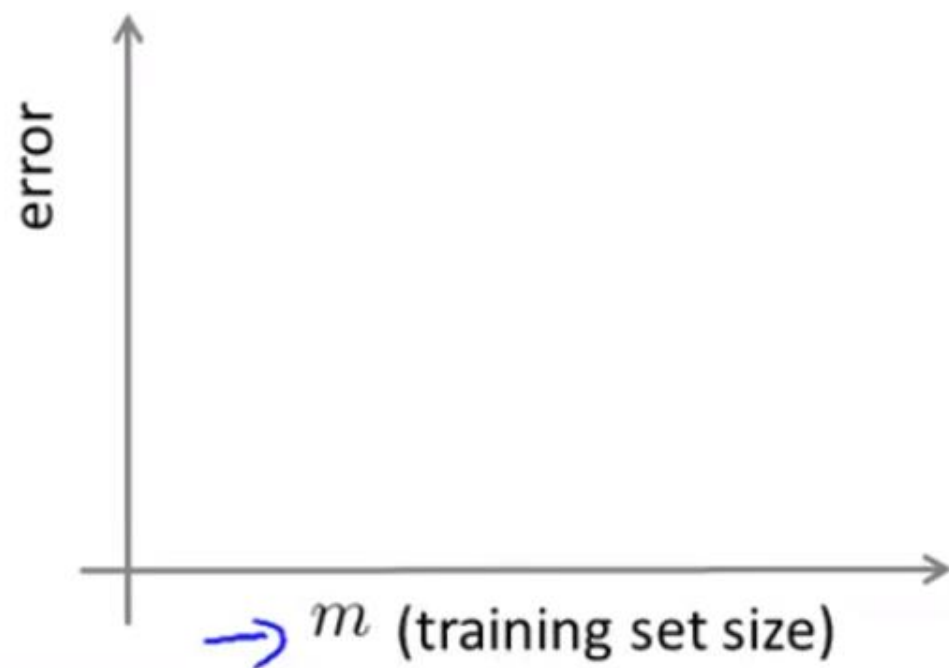


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

## Learning curves

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

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

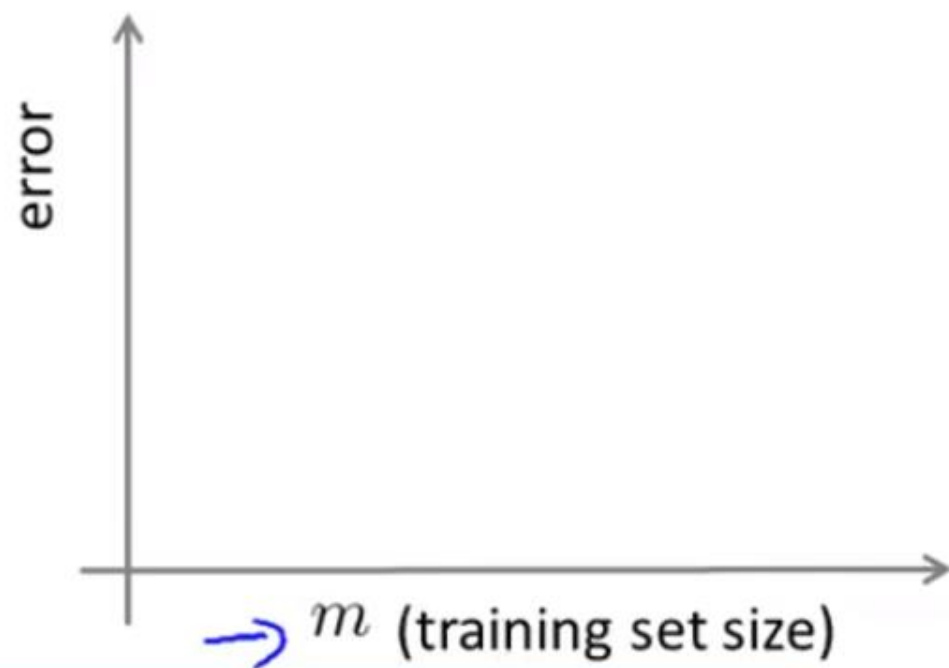


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

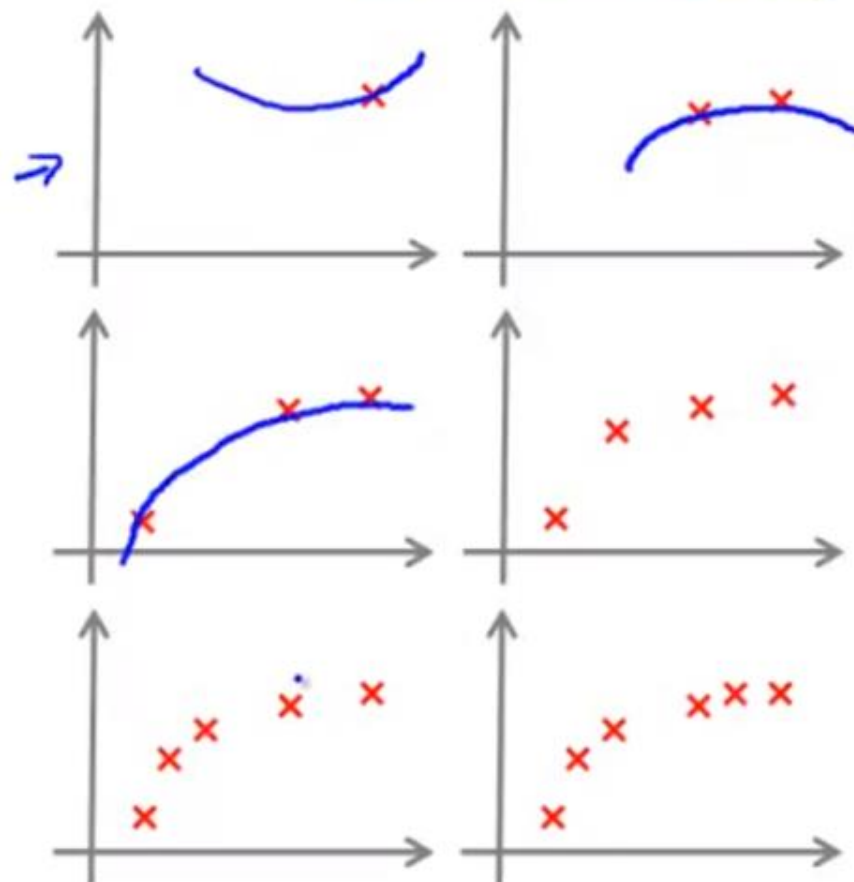
# Learning curves

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

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

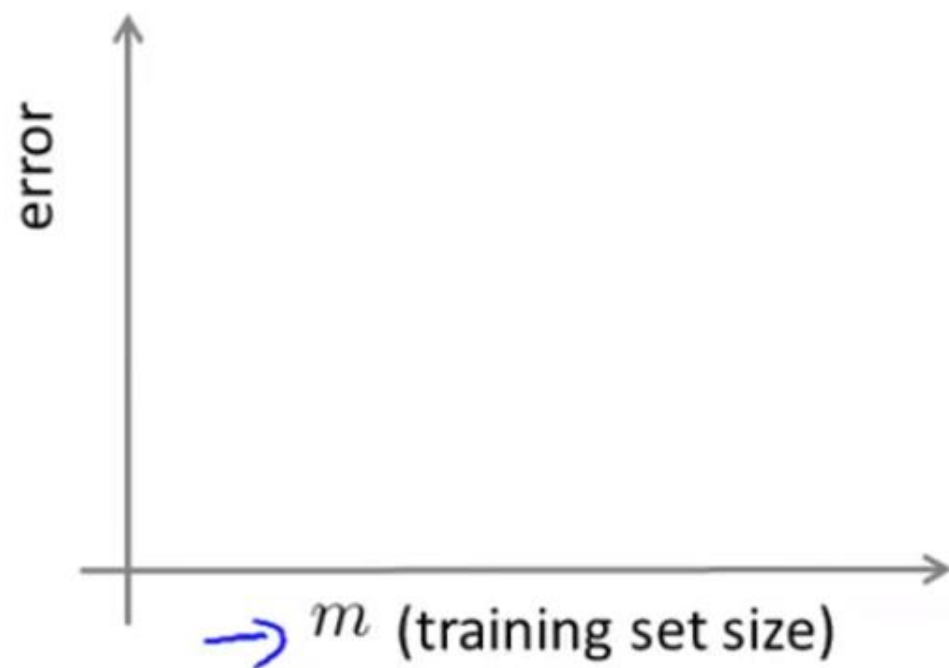


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

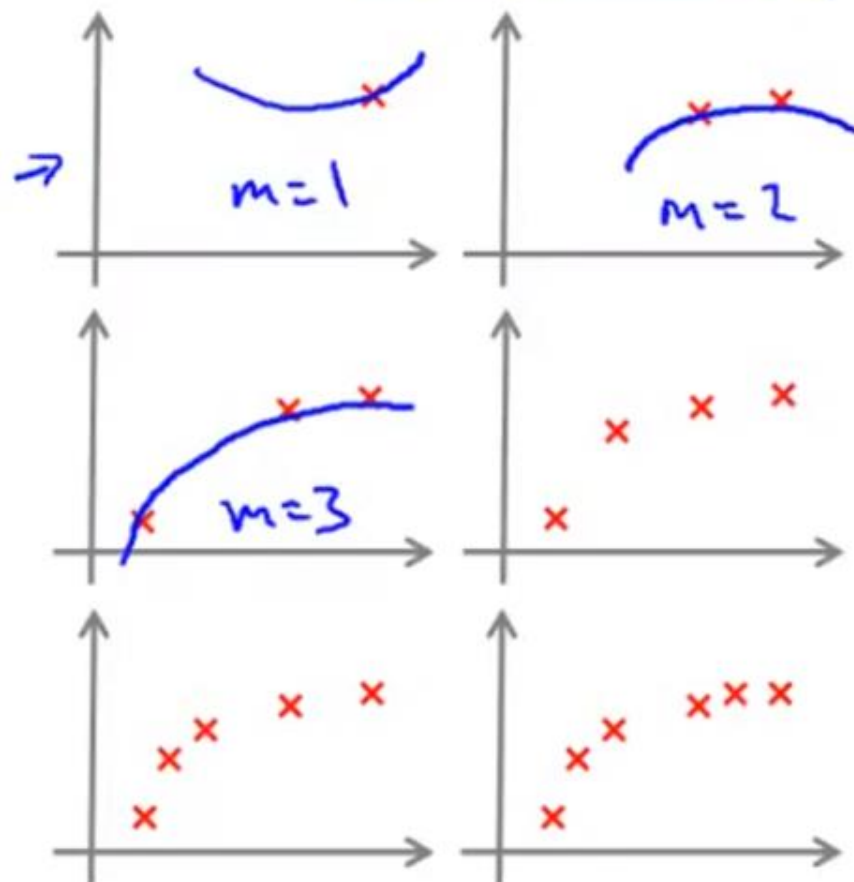
# Learning curves

$$\rightarrow J_{train}(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



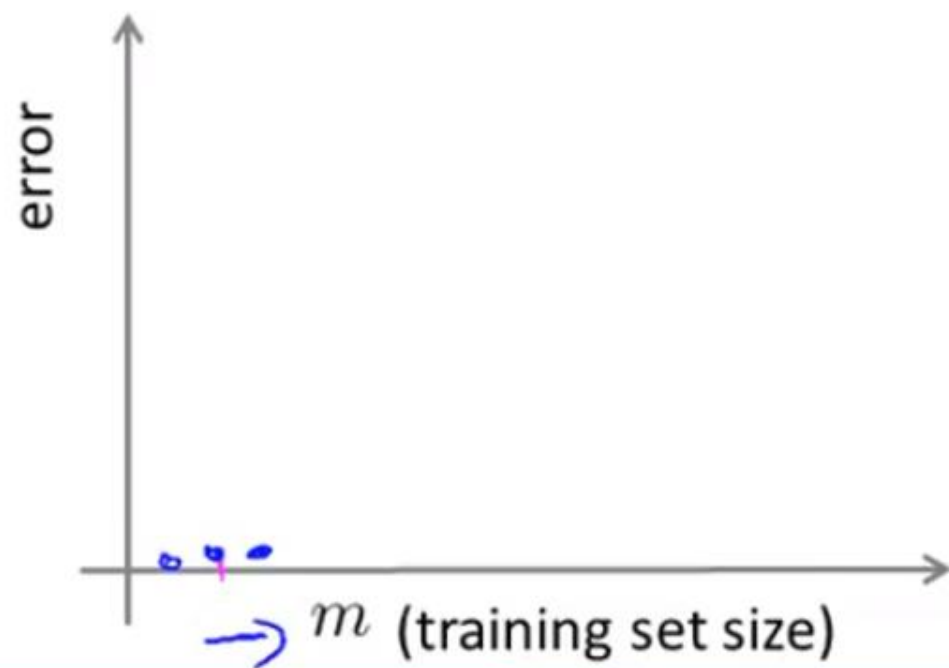
Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.



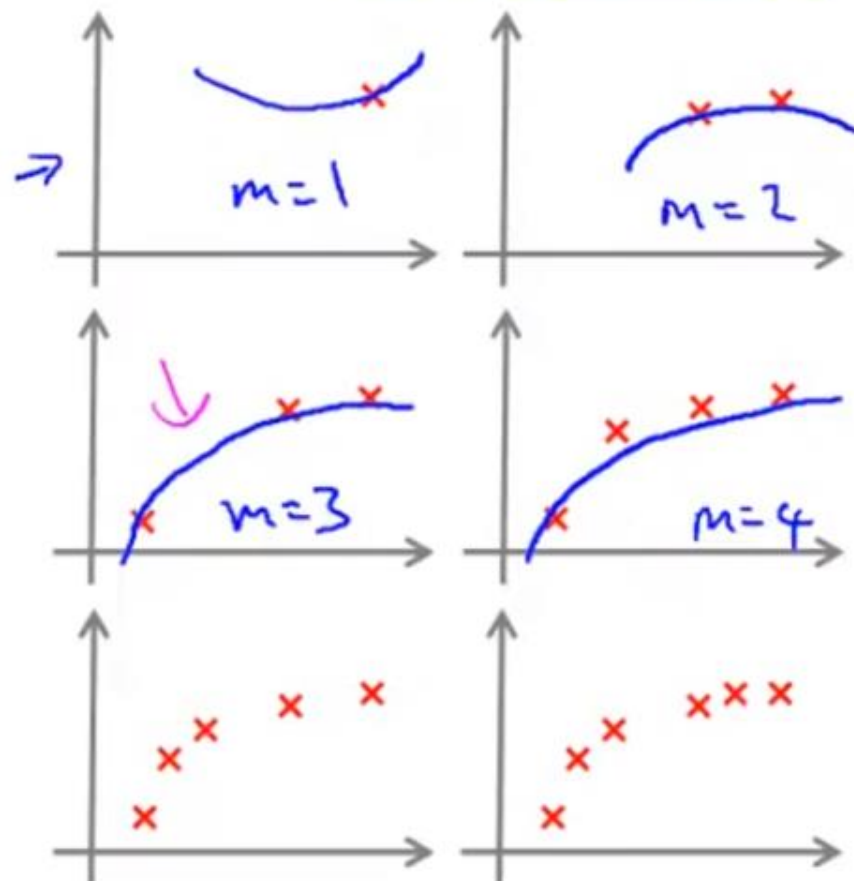
# Learning curves

$$\rightarrow \underline{J_{train}(\theta)} = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

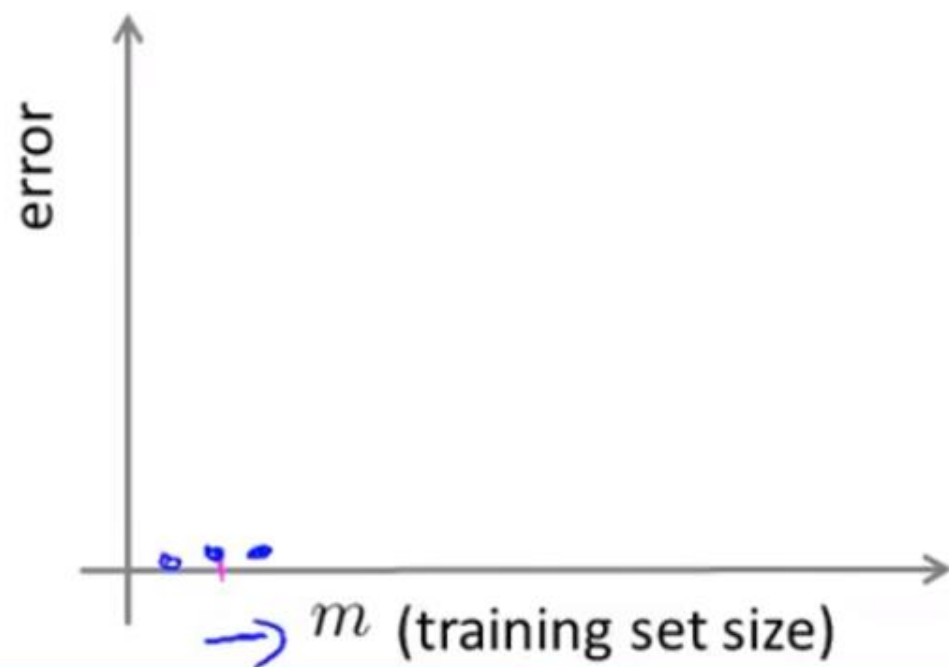


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

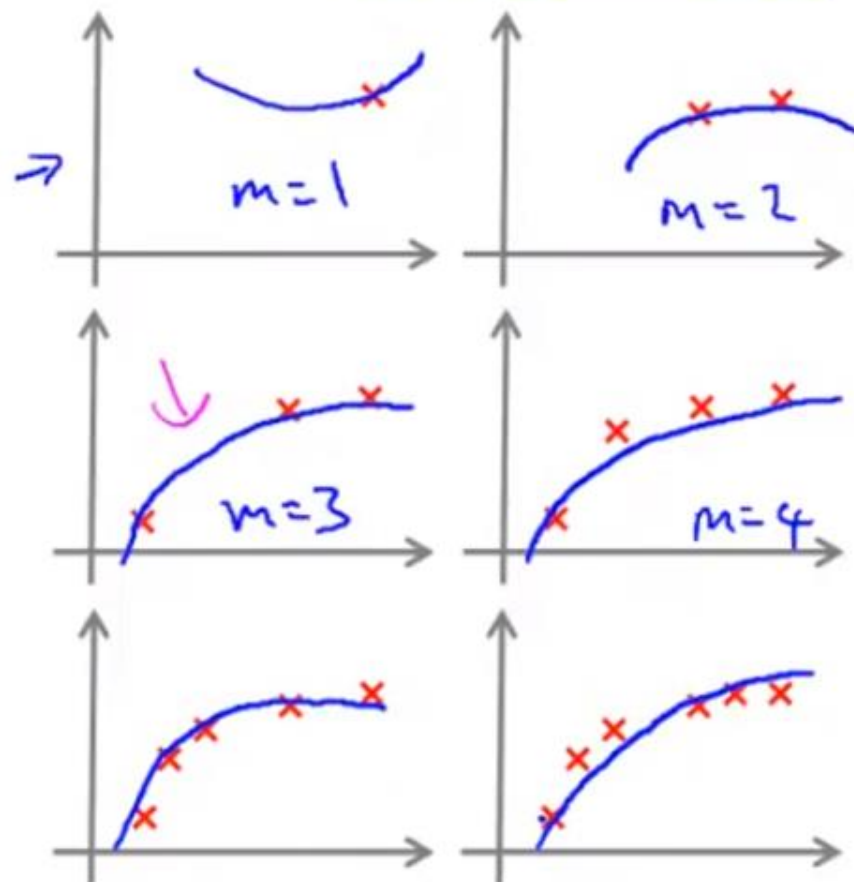
# Learning curves

$$\rightarrow \underline{J_{train}(\theta)} = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

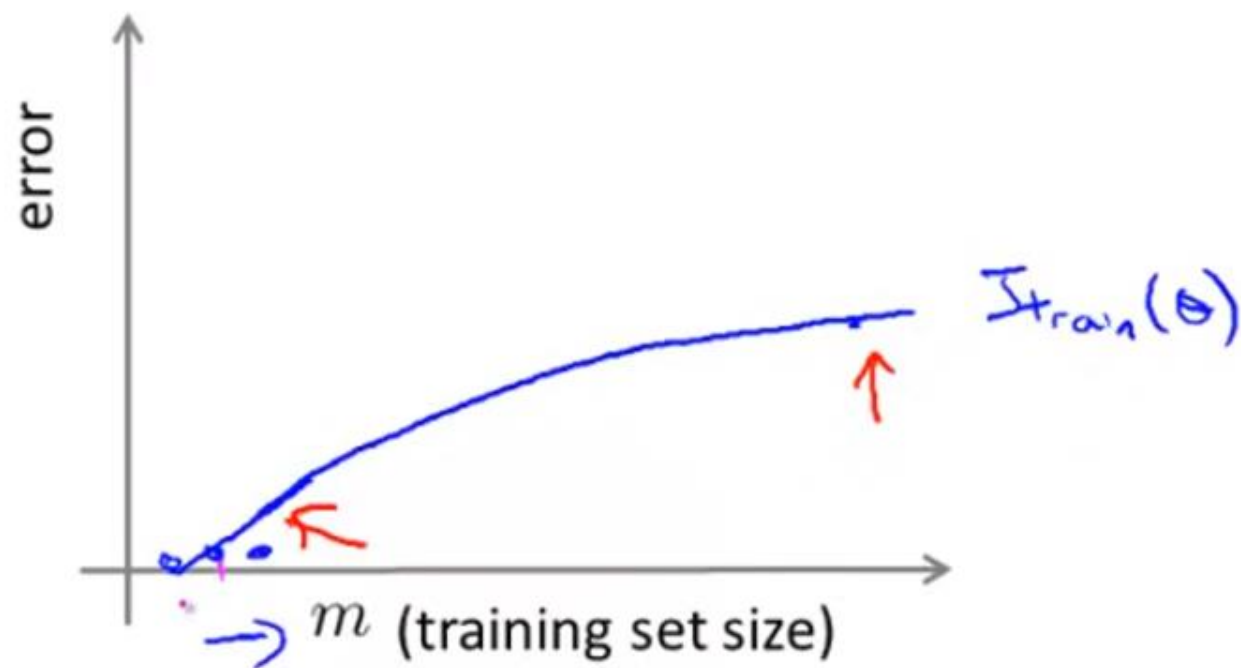


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

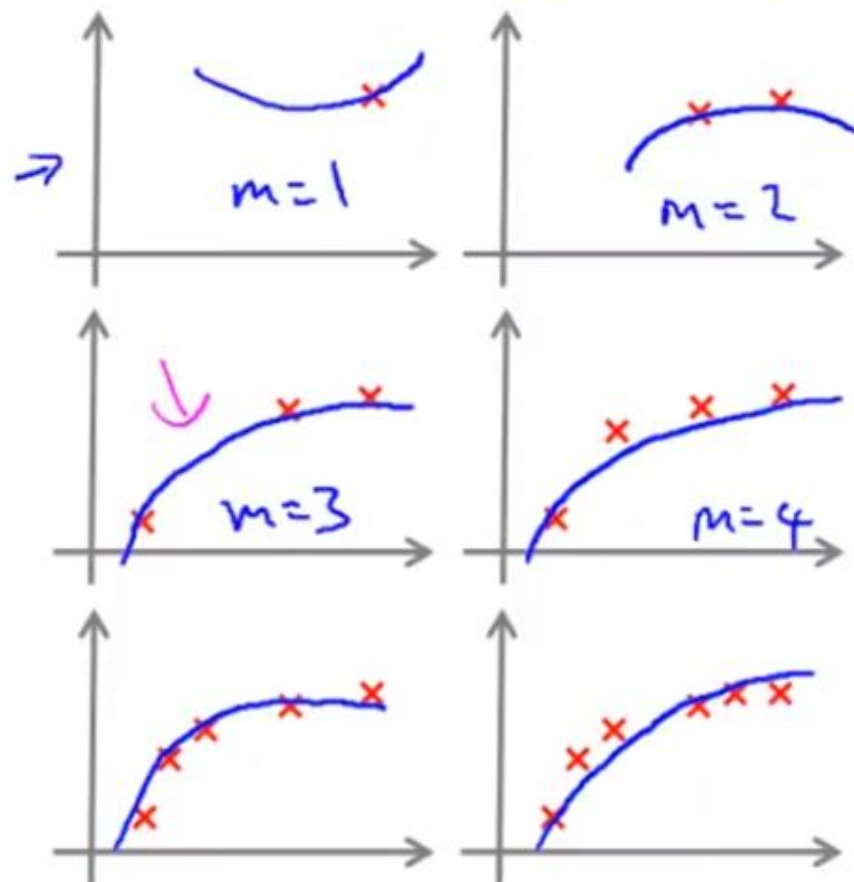
# Learning curves

$$\rightarrow \underline{J_{train}(\theta)} = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

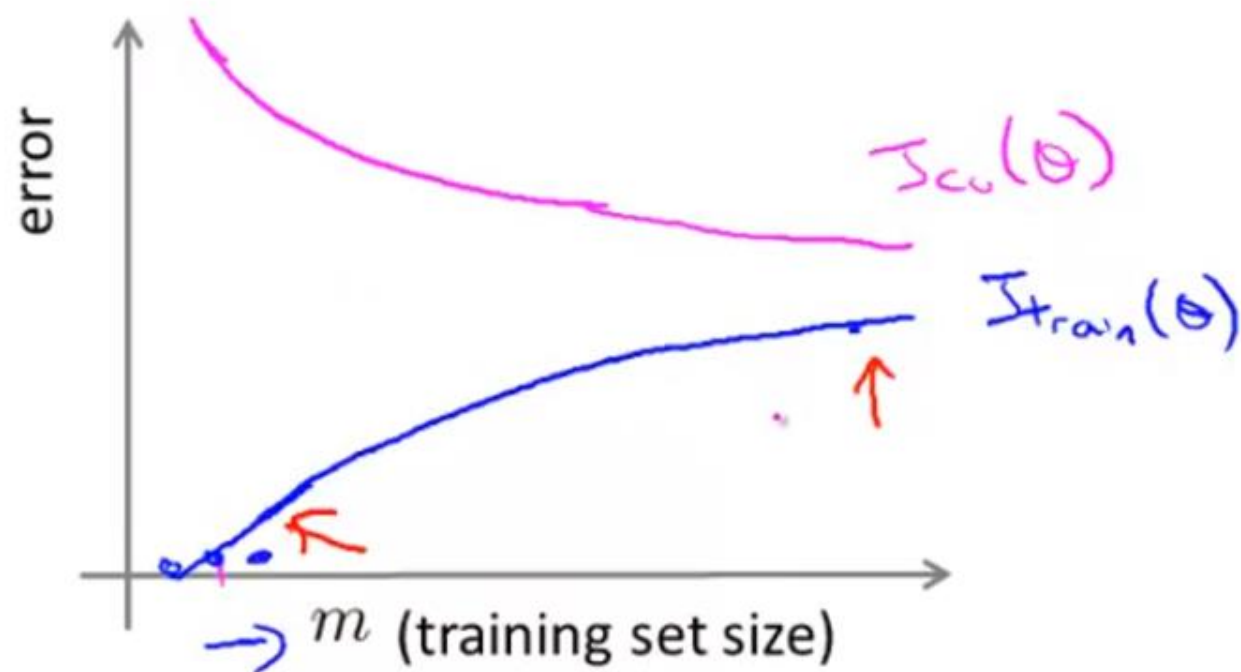


Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

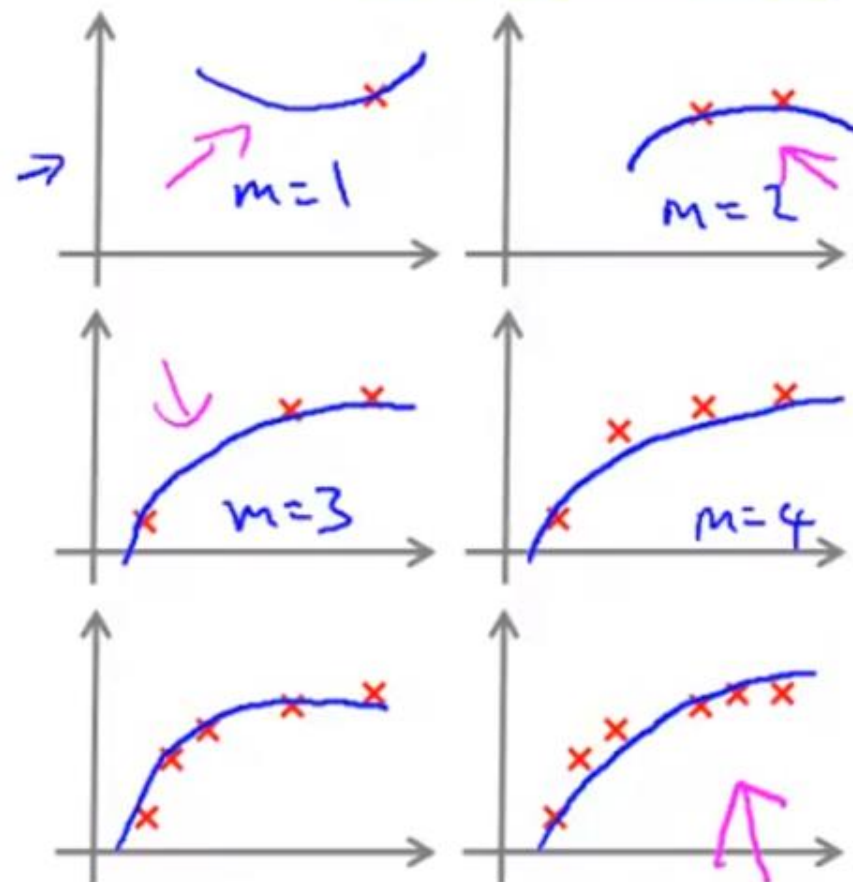
# Learning curves

$$\rightarrow \underline{J_{train}(\theta)} = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \leftarrow$$

$$\rightarrow J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

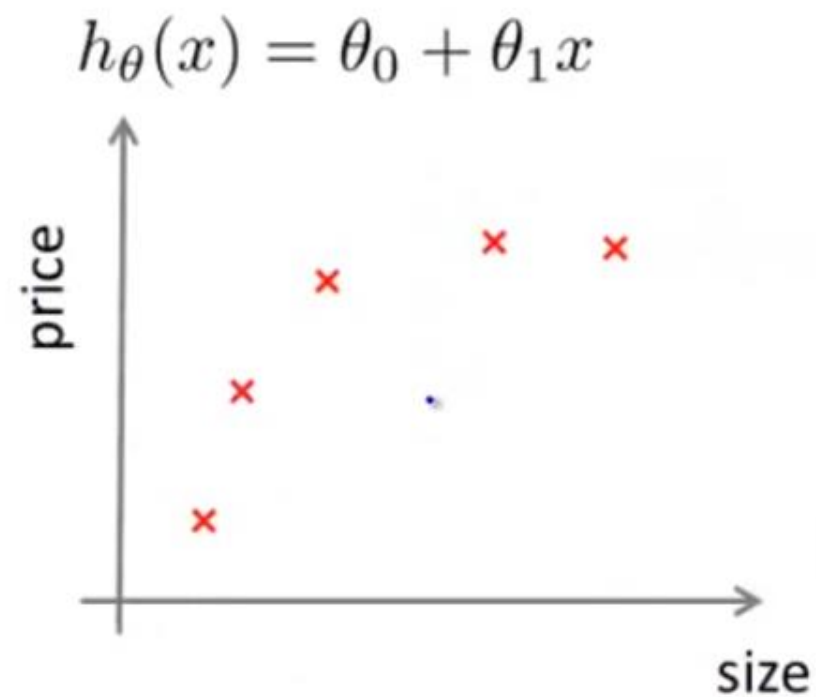
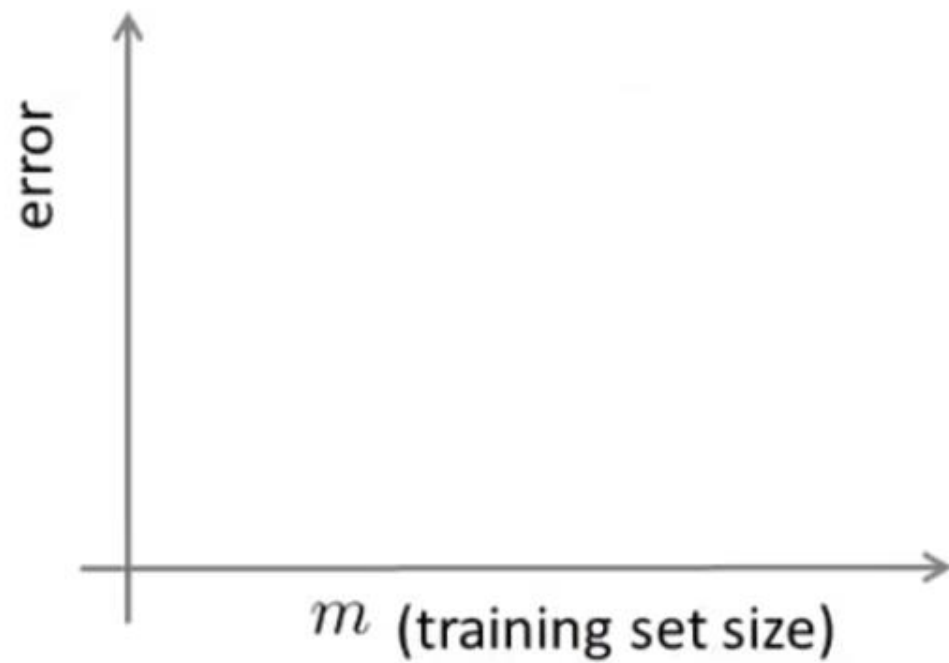


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



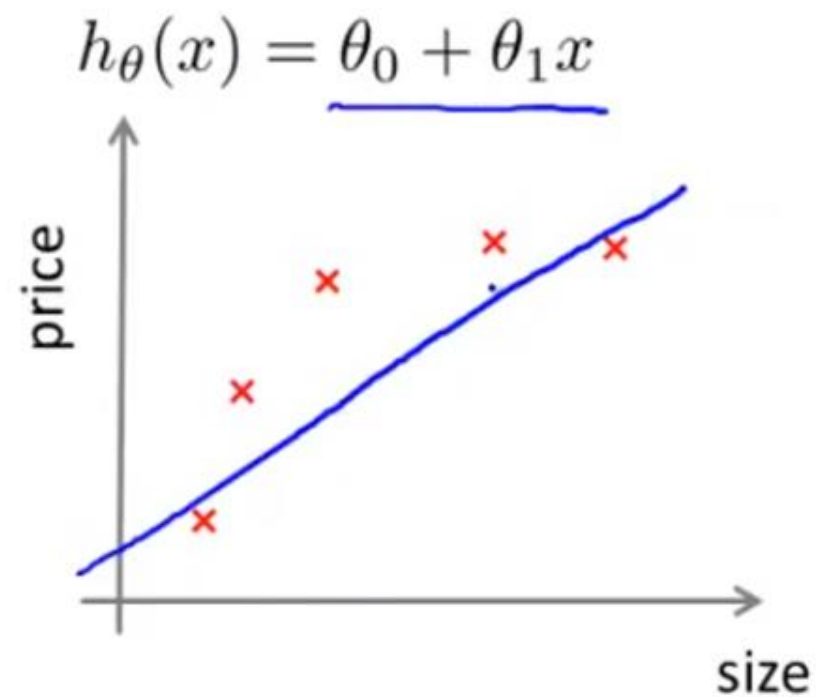
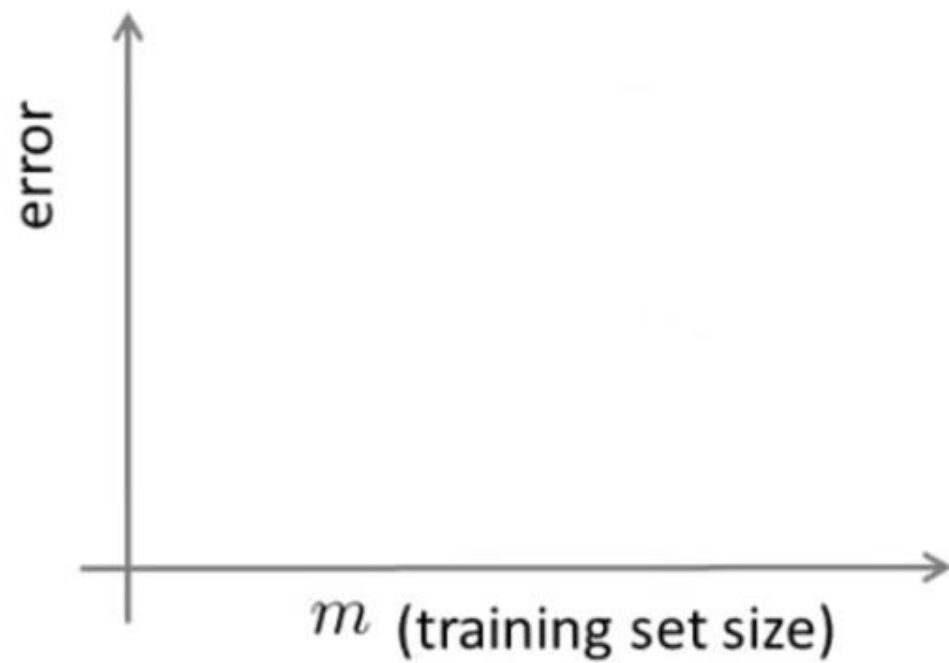
Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

## High bias



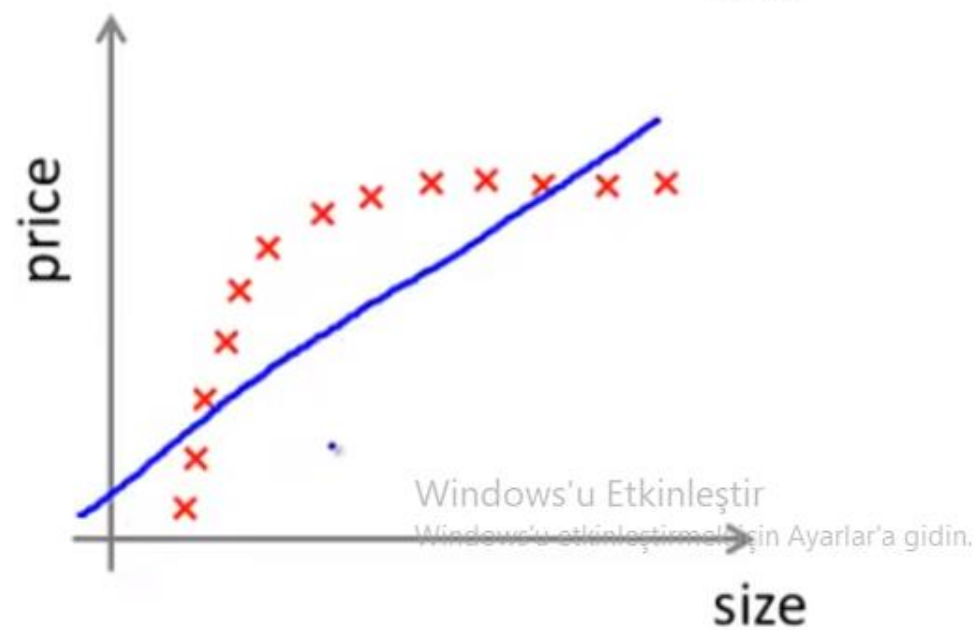
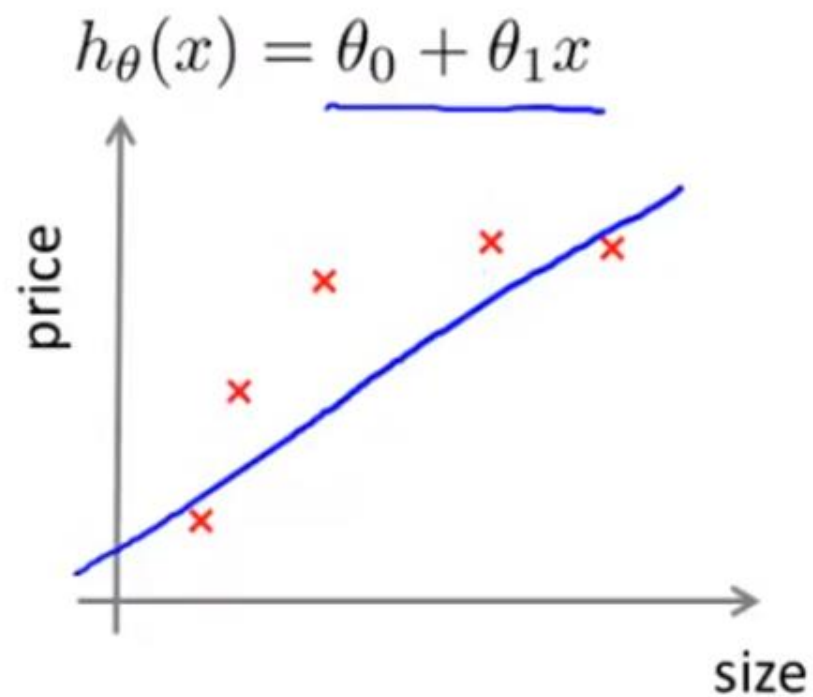
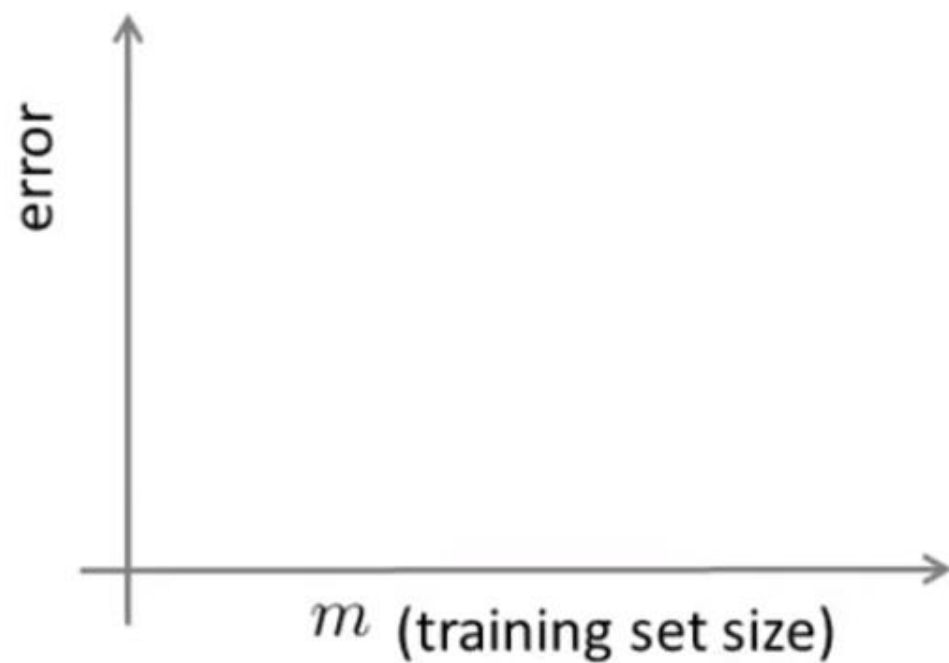
Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

## High bias



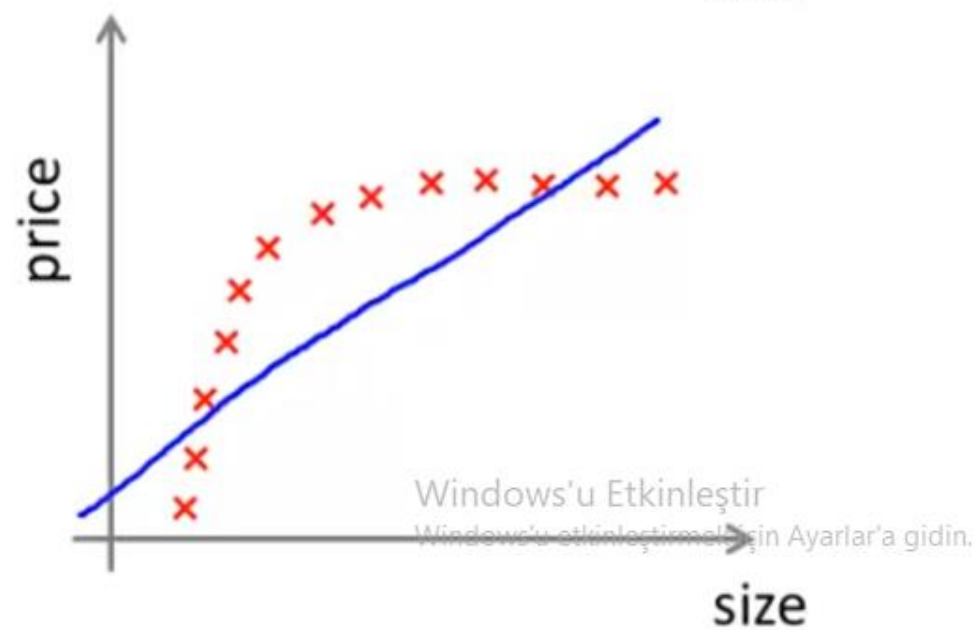
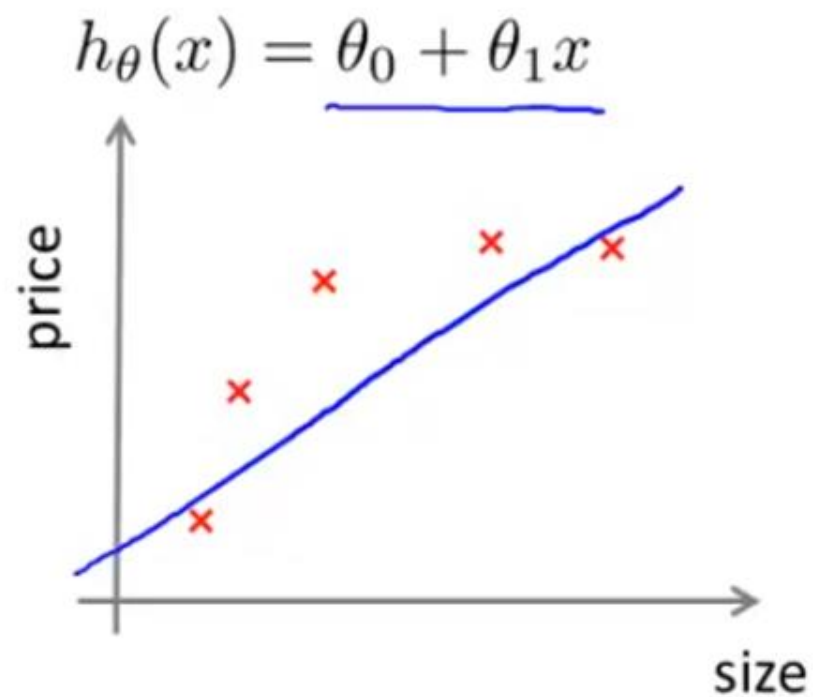
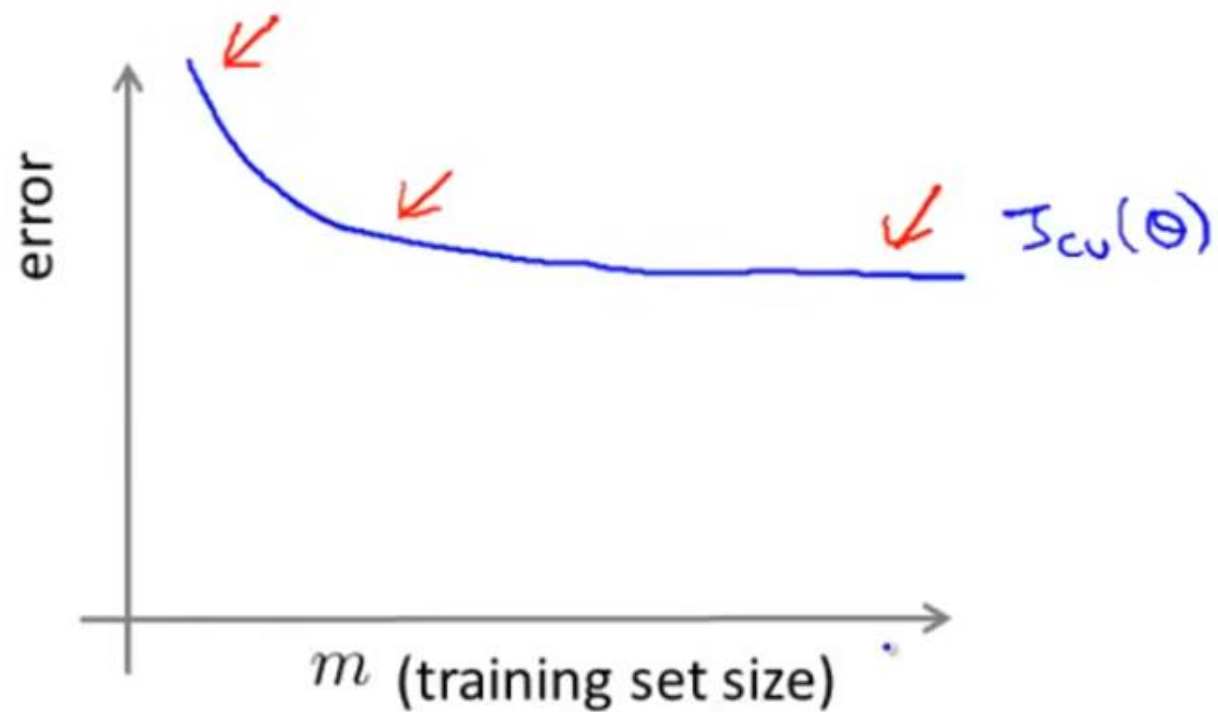
Windows'u Etkinleştir  
Windows'u etkinleştirmek için Ayarlar'a gidin.

## High bias



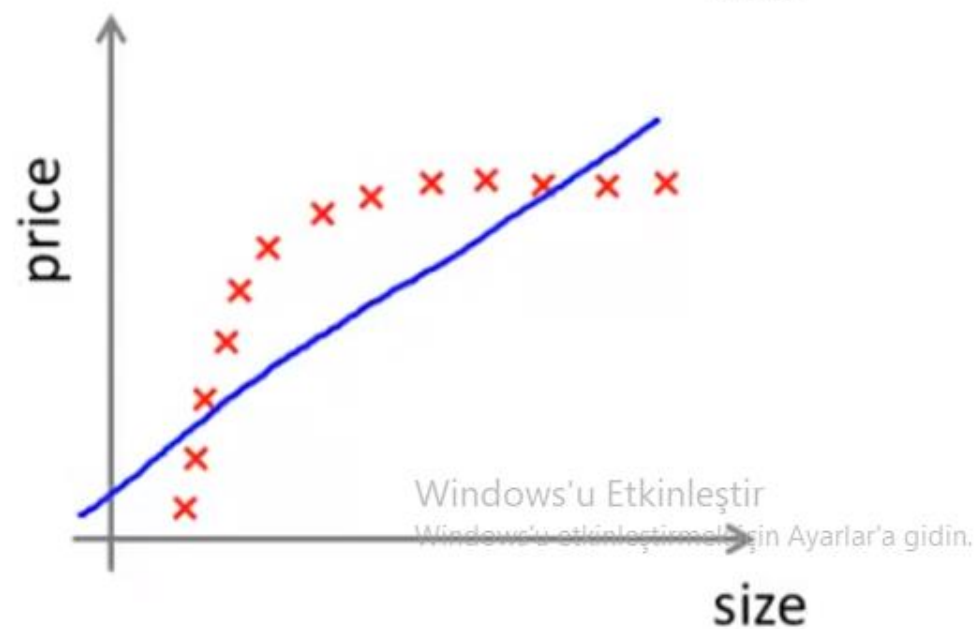
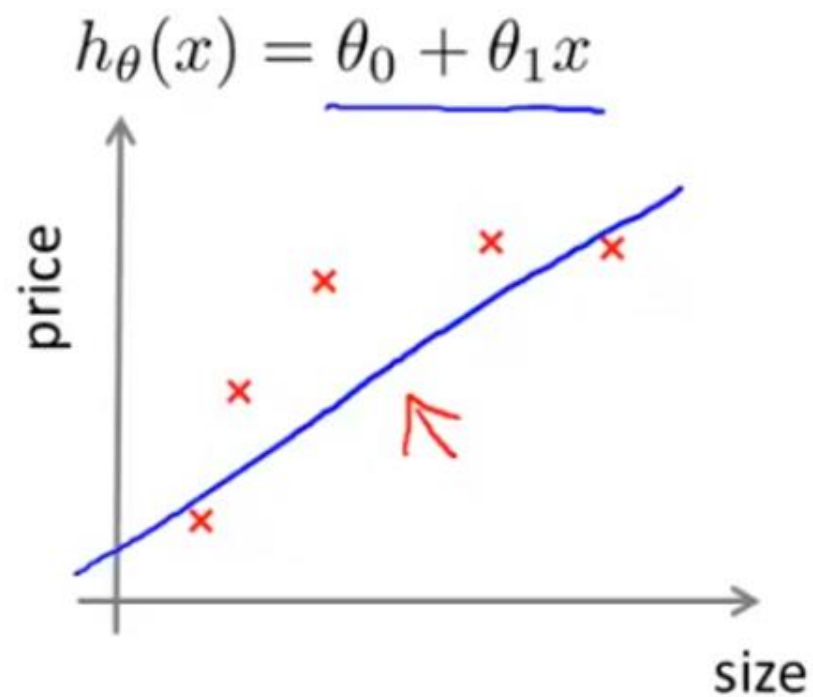
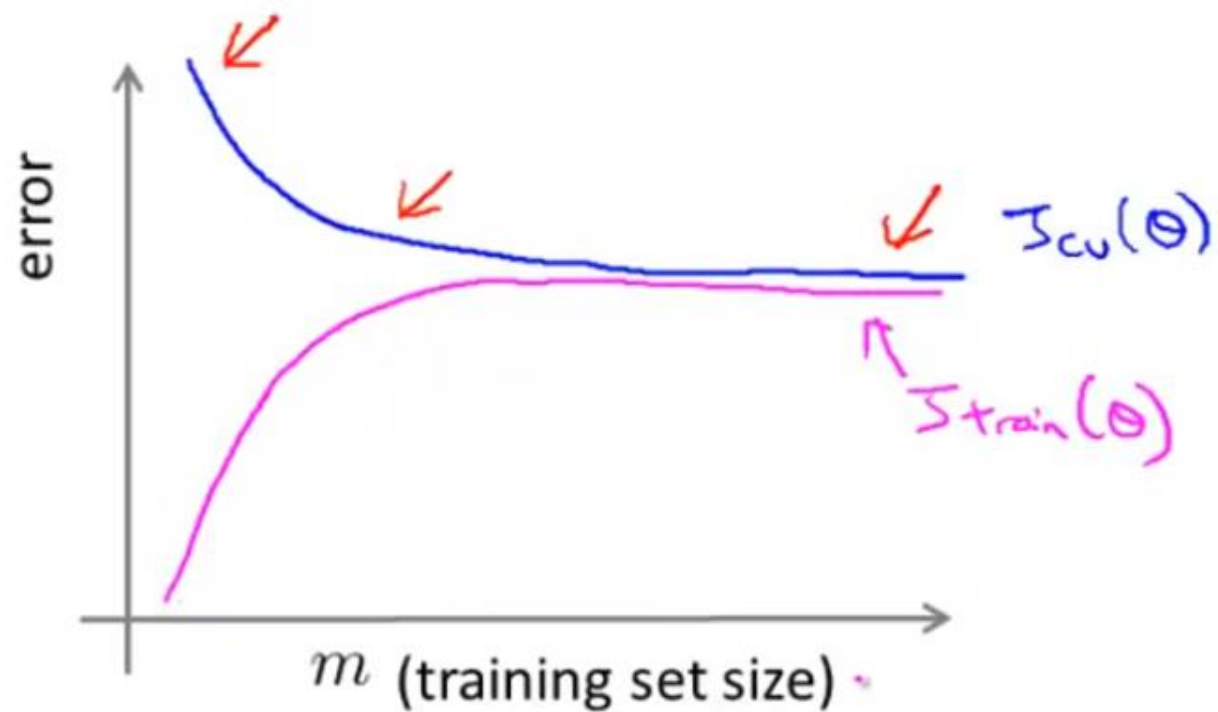


## High bias

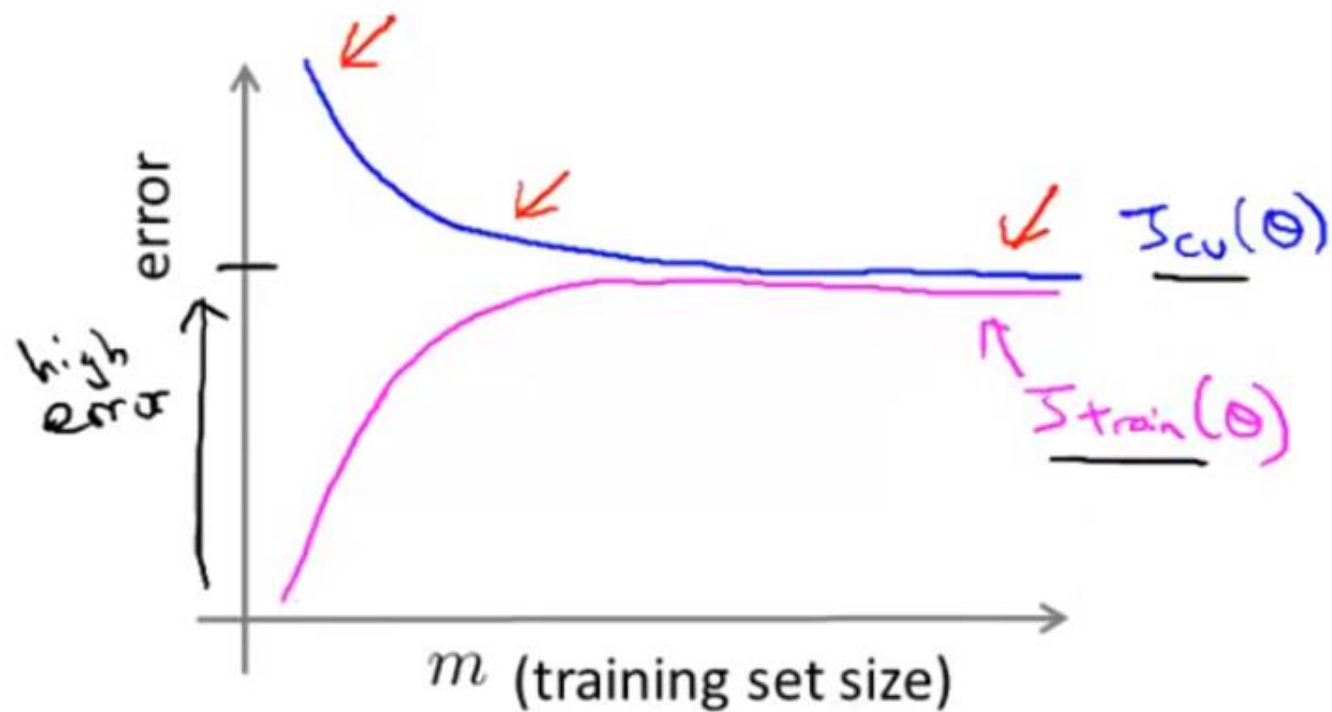




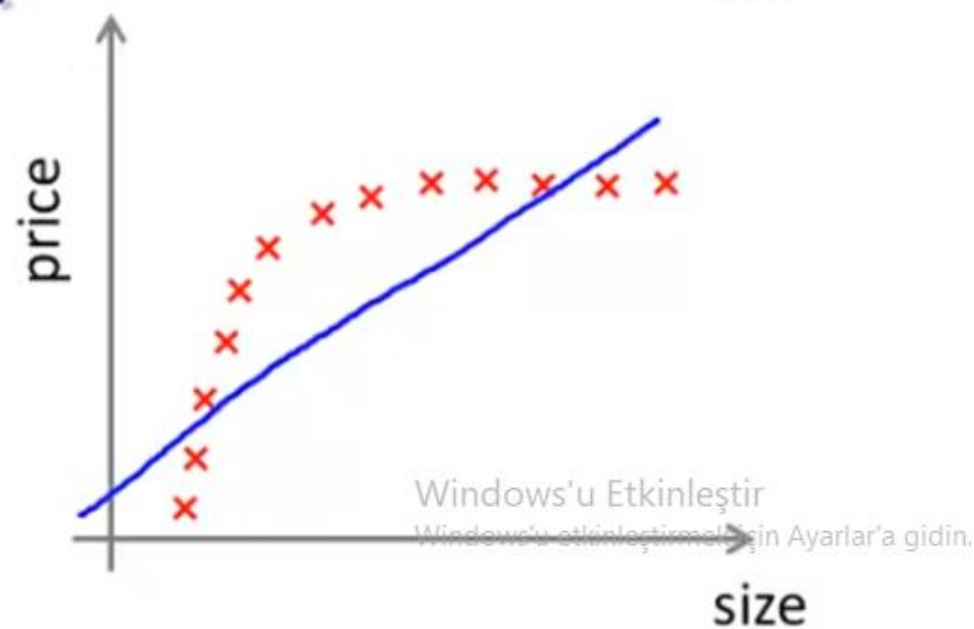
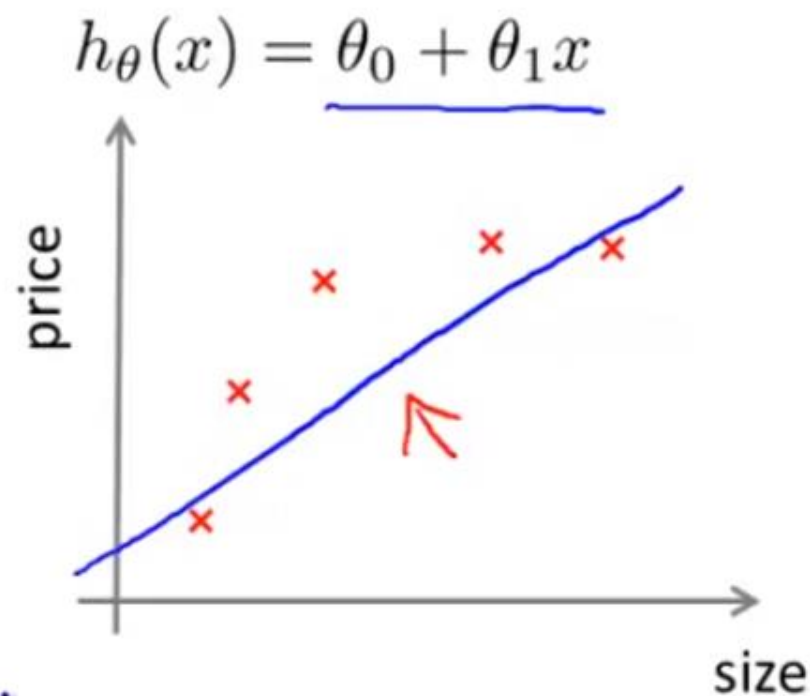
## High bias



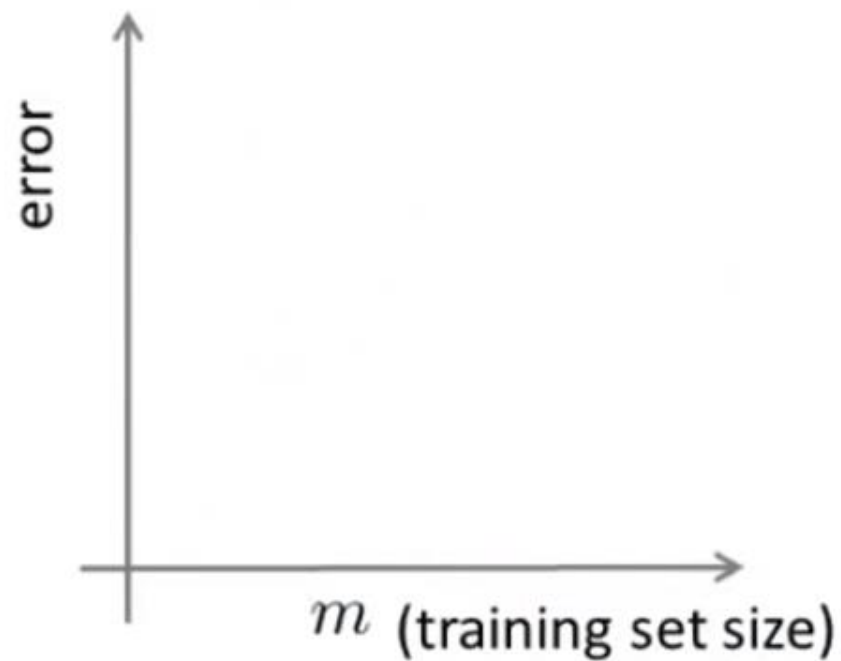
## High bias



If a learning algorithm is suffering from high bias, getting more training data will not (by itself) help much.

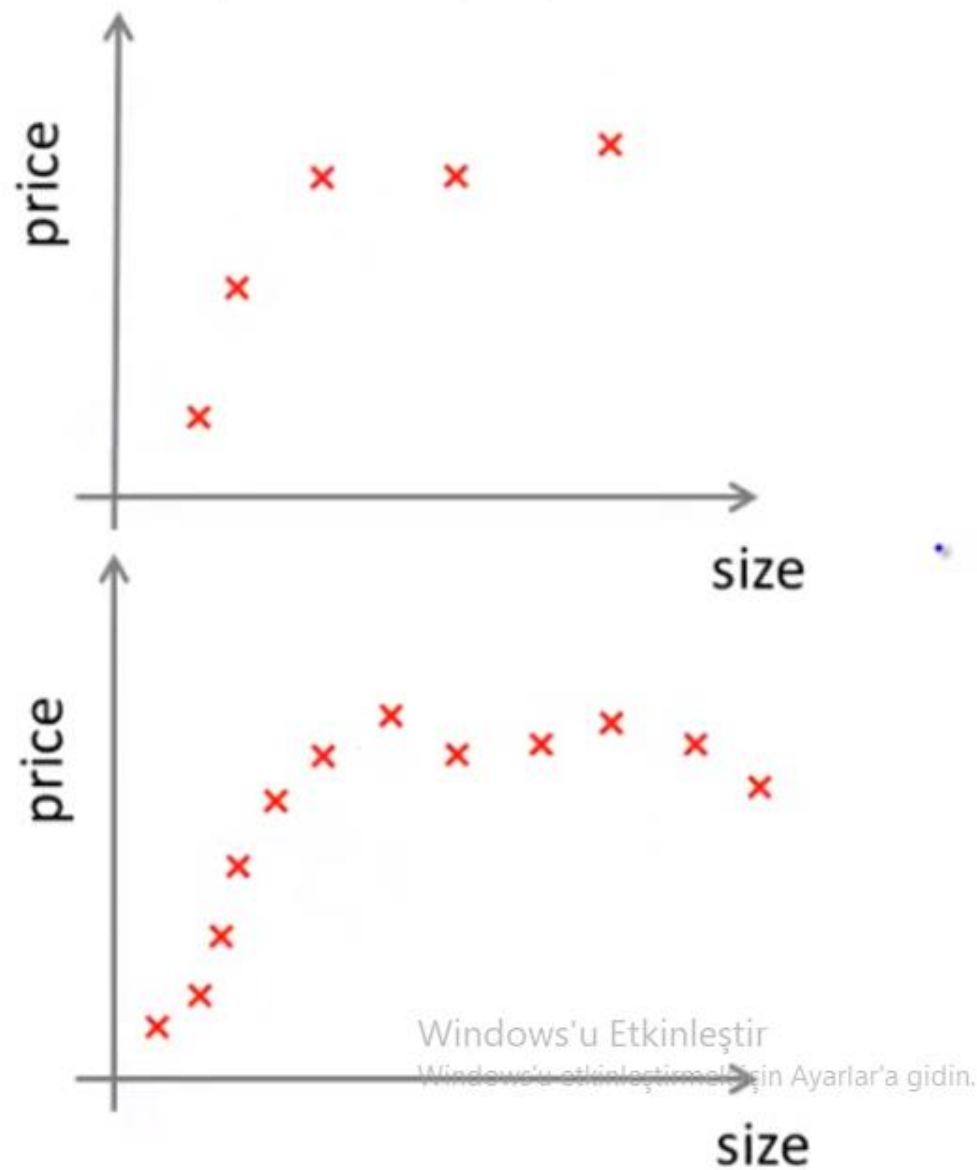


## High variance

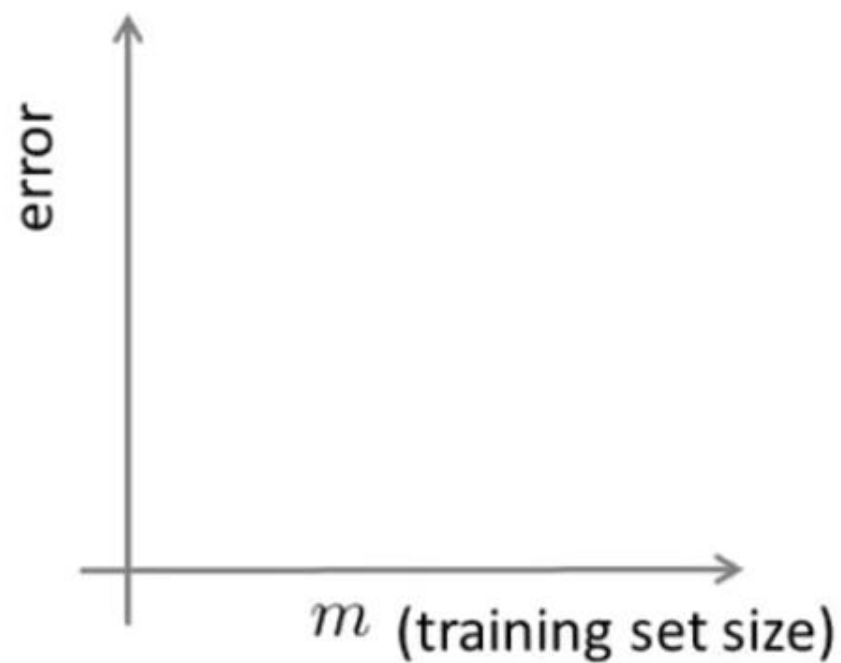


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )

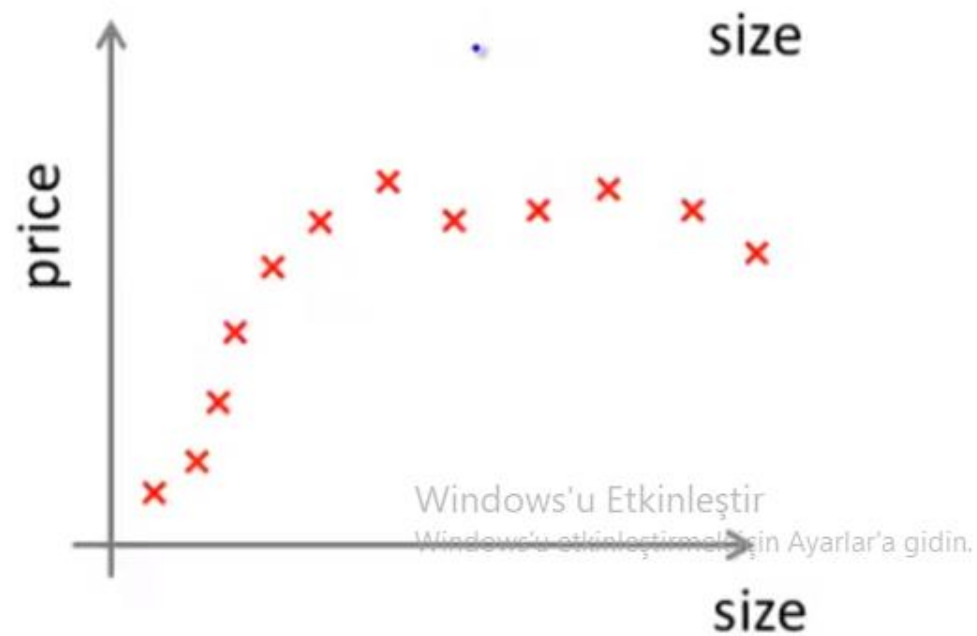
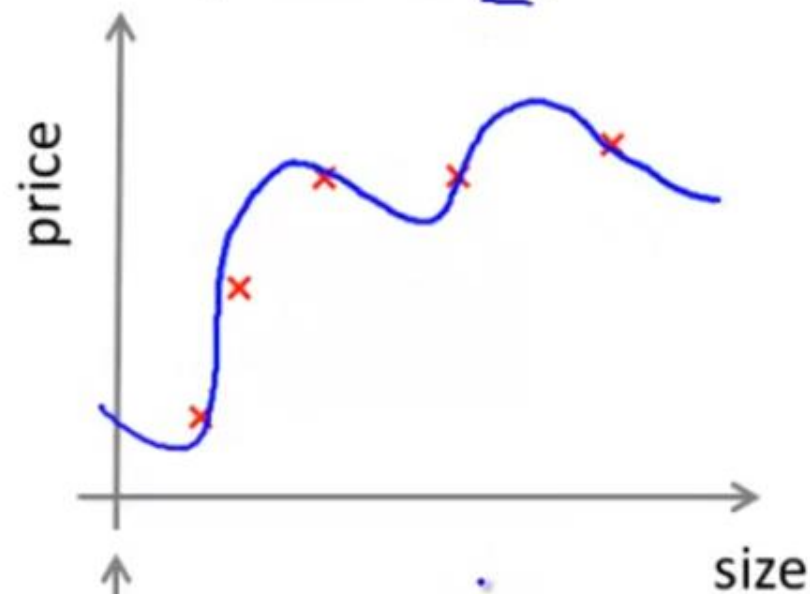


## High variance

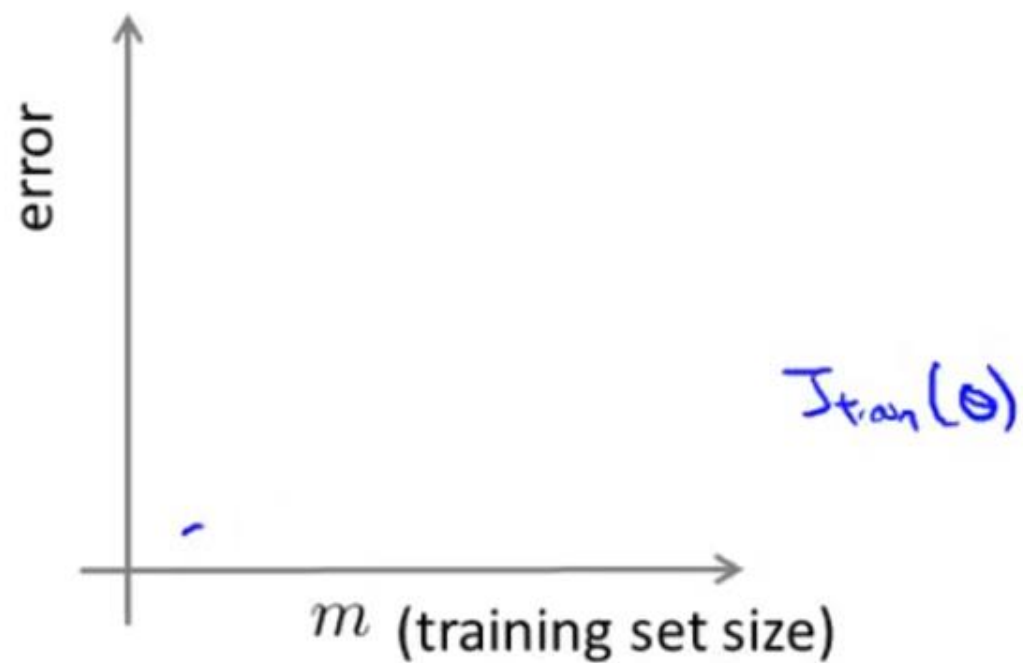


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )

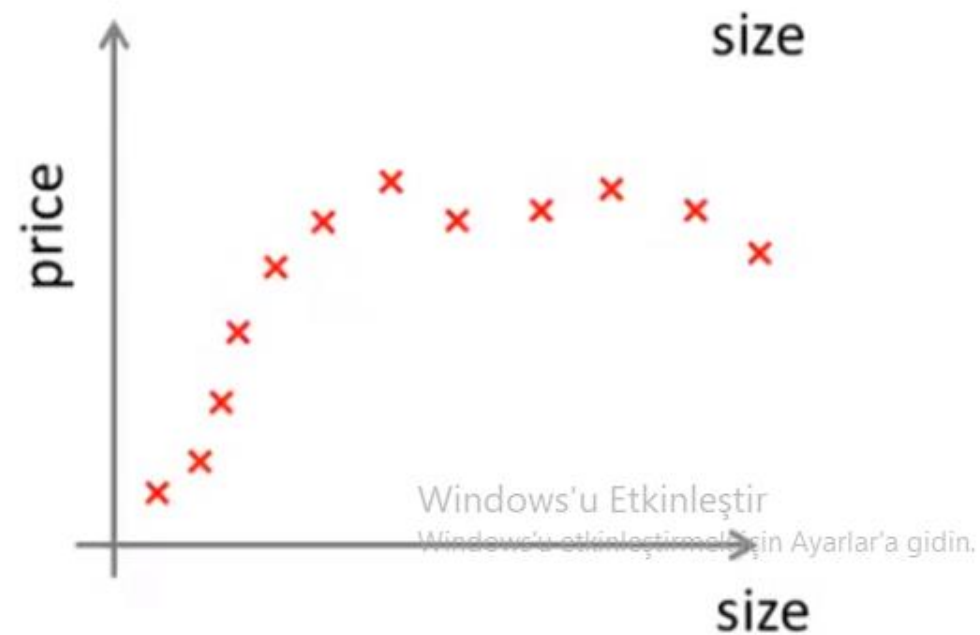
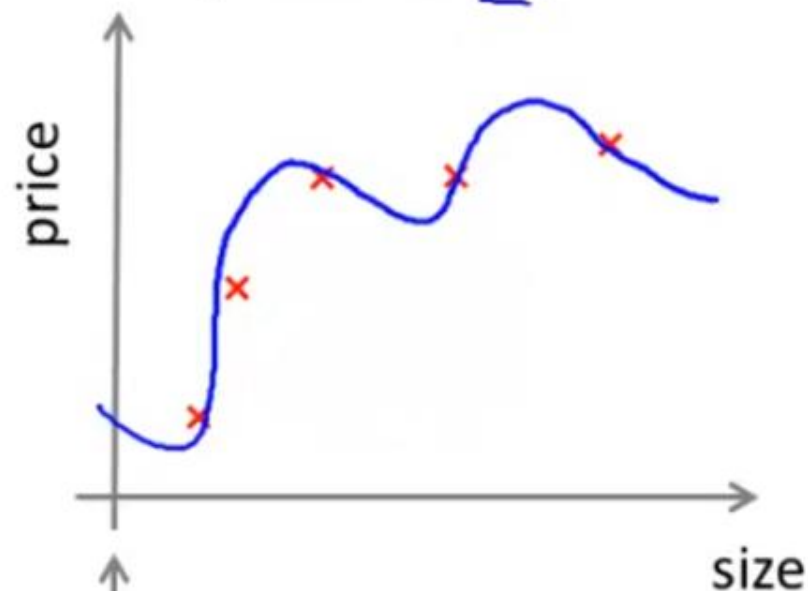


## High variance

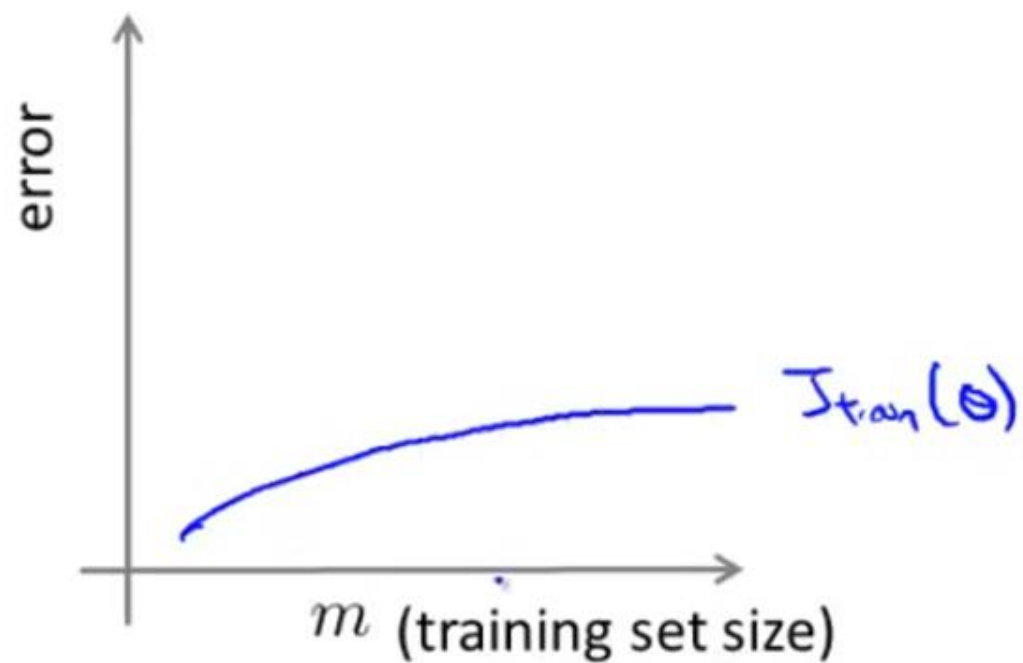


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )

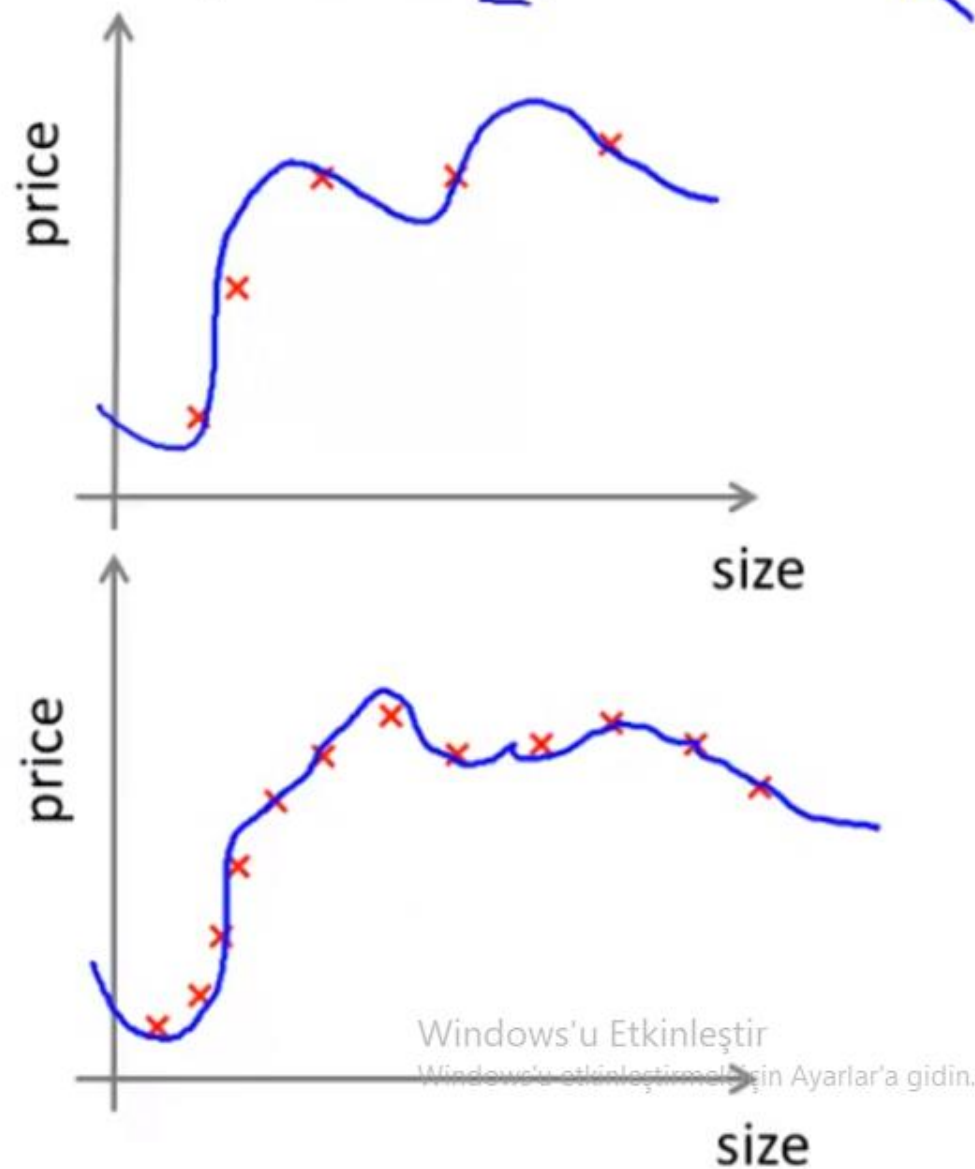


## High variance

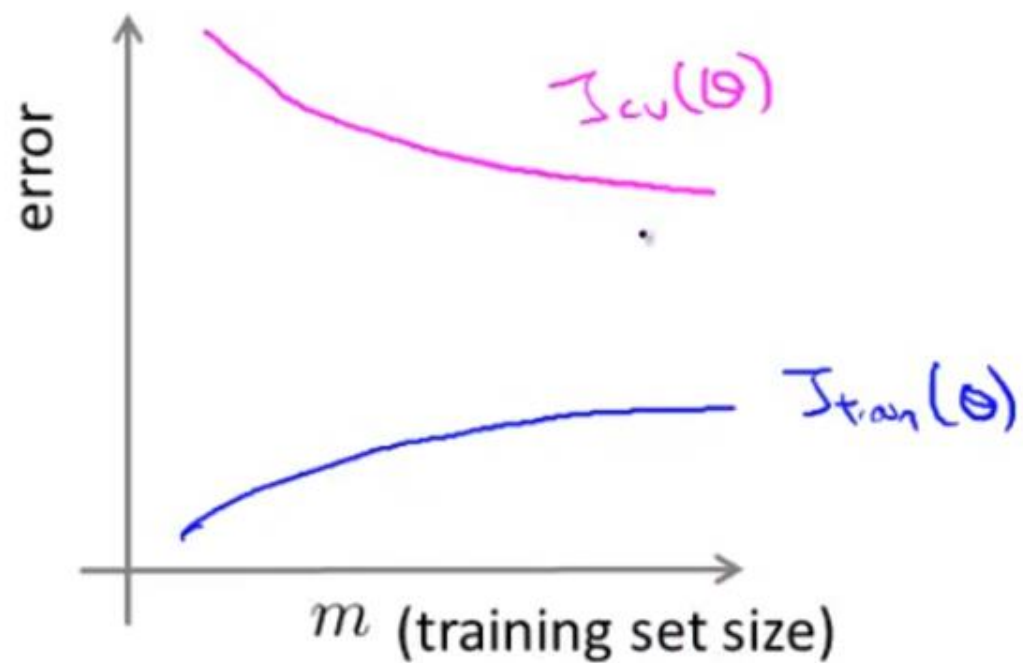


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )

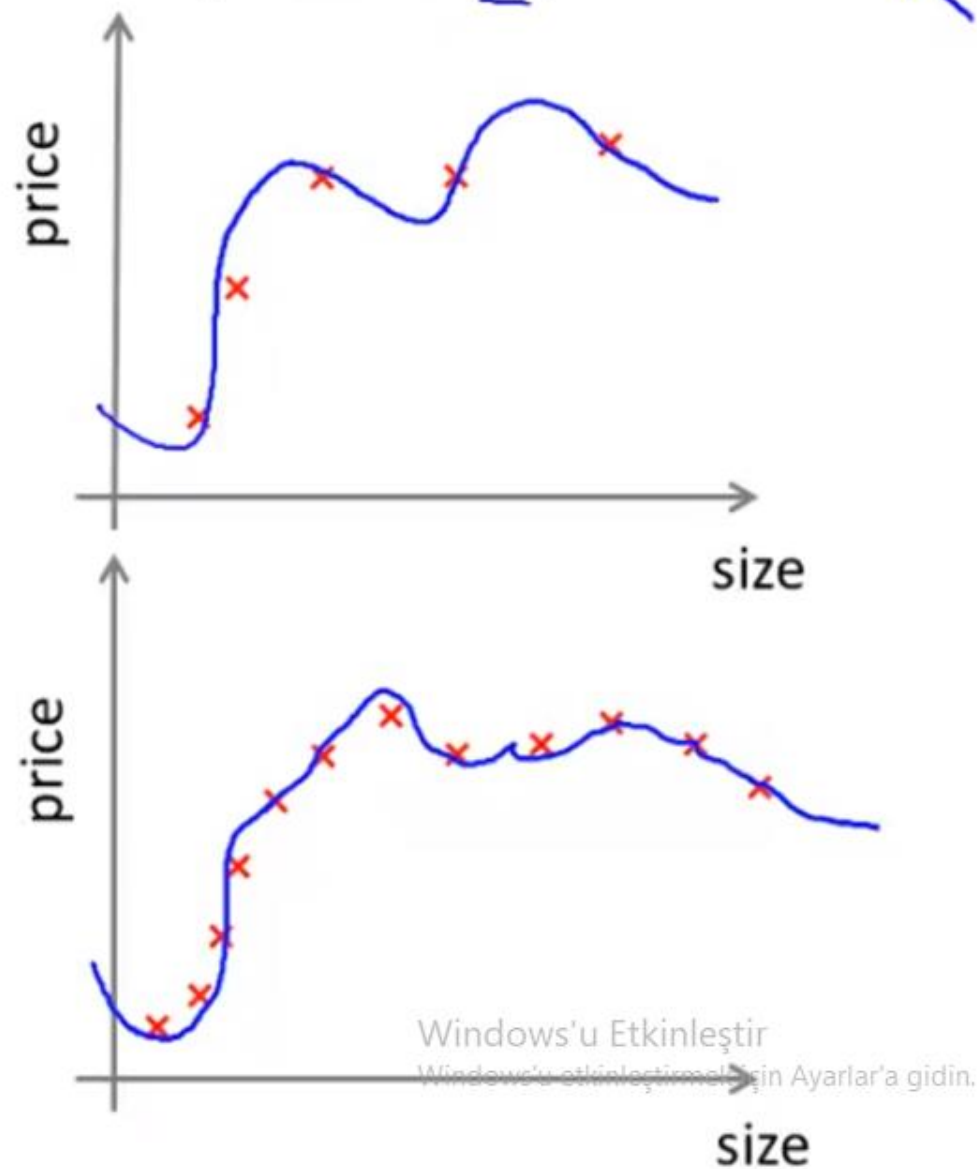


## High variance

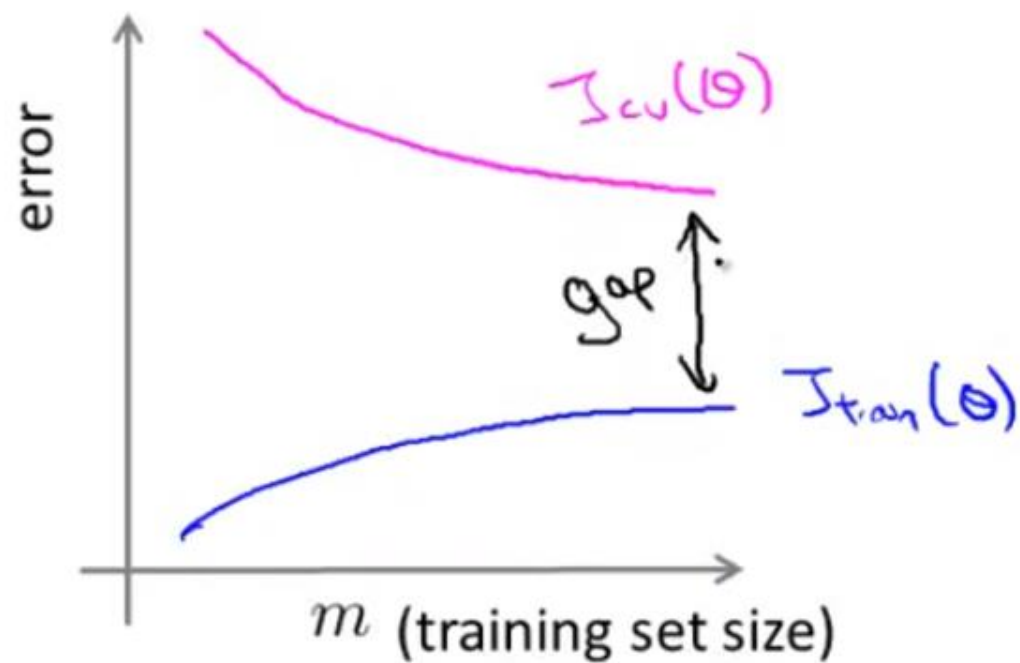


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )

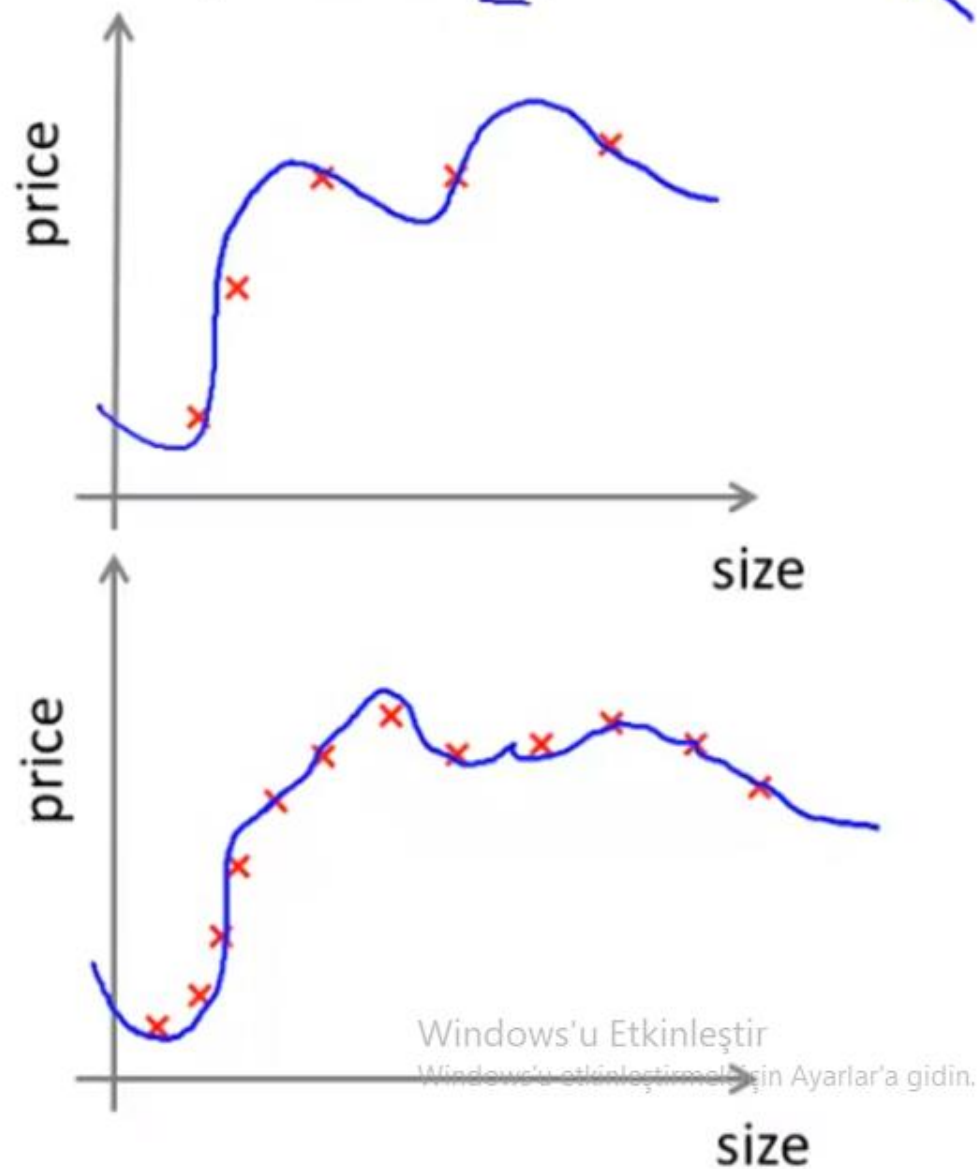


## High variance



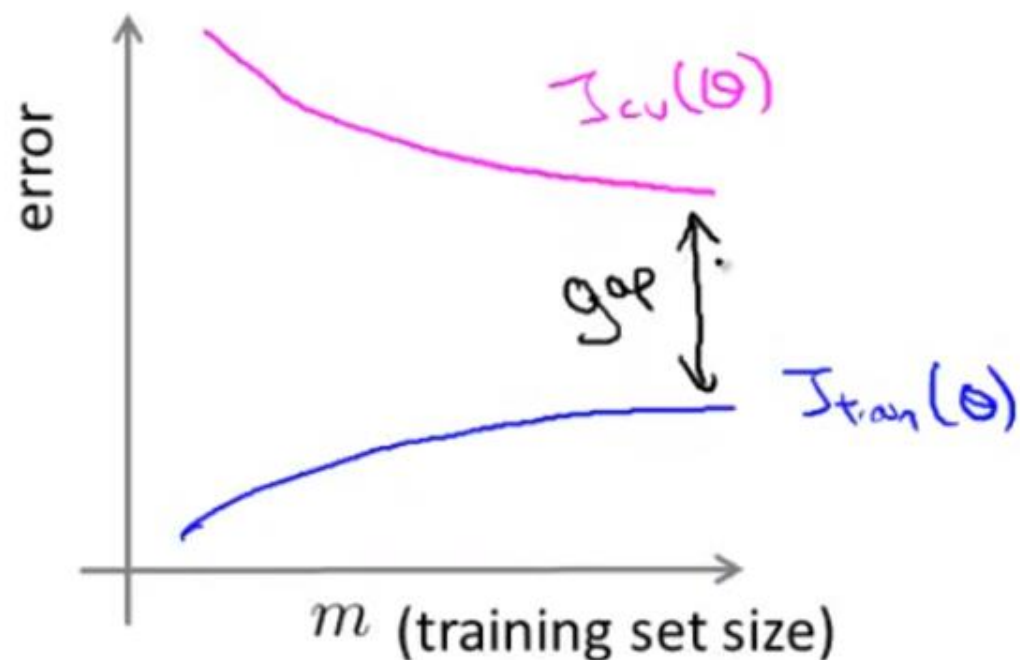
$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )





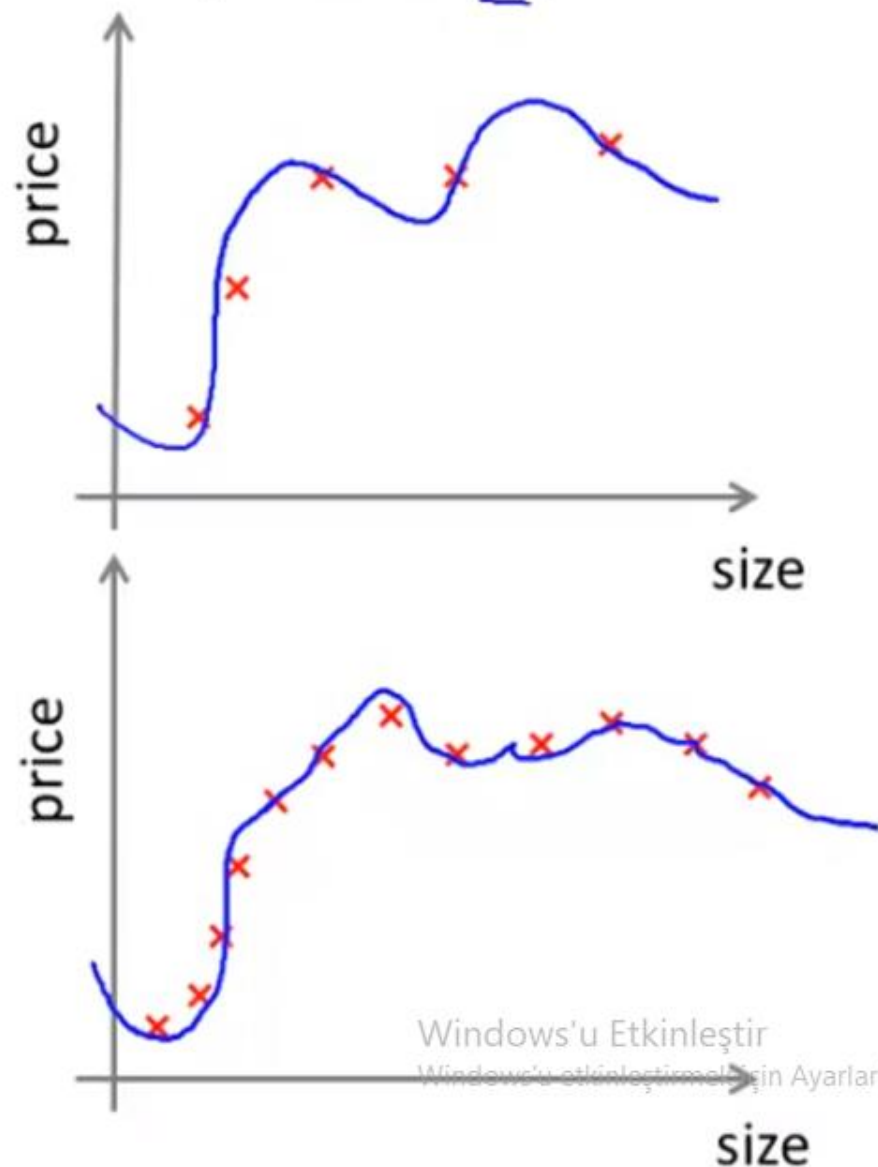
## High variance



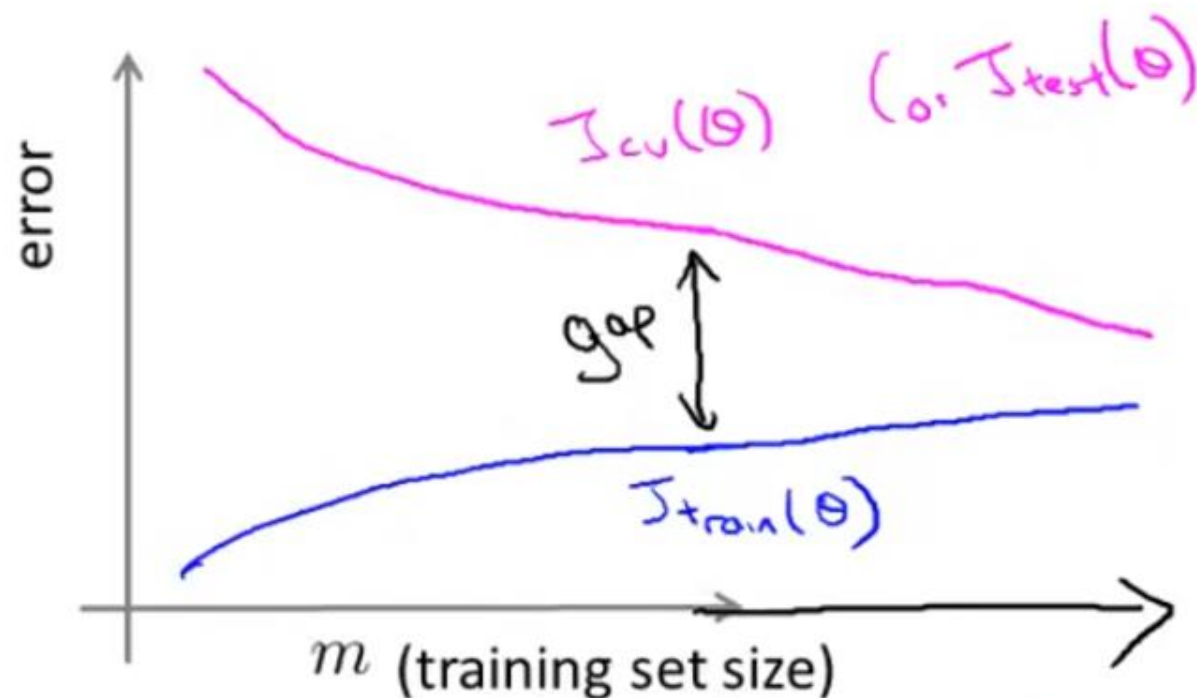
If a learning algorithm is suffering from high variance, getting more training data is likely to help.

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )



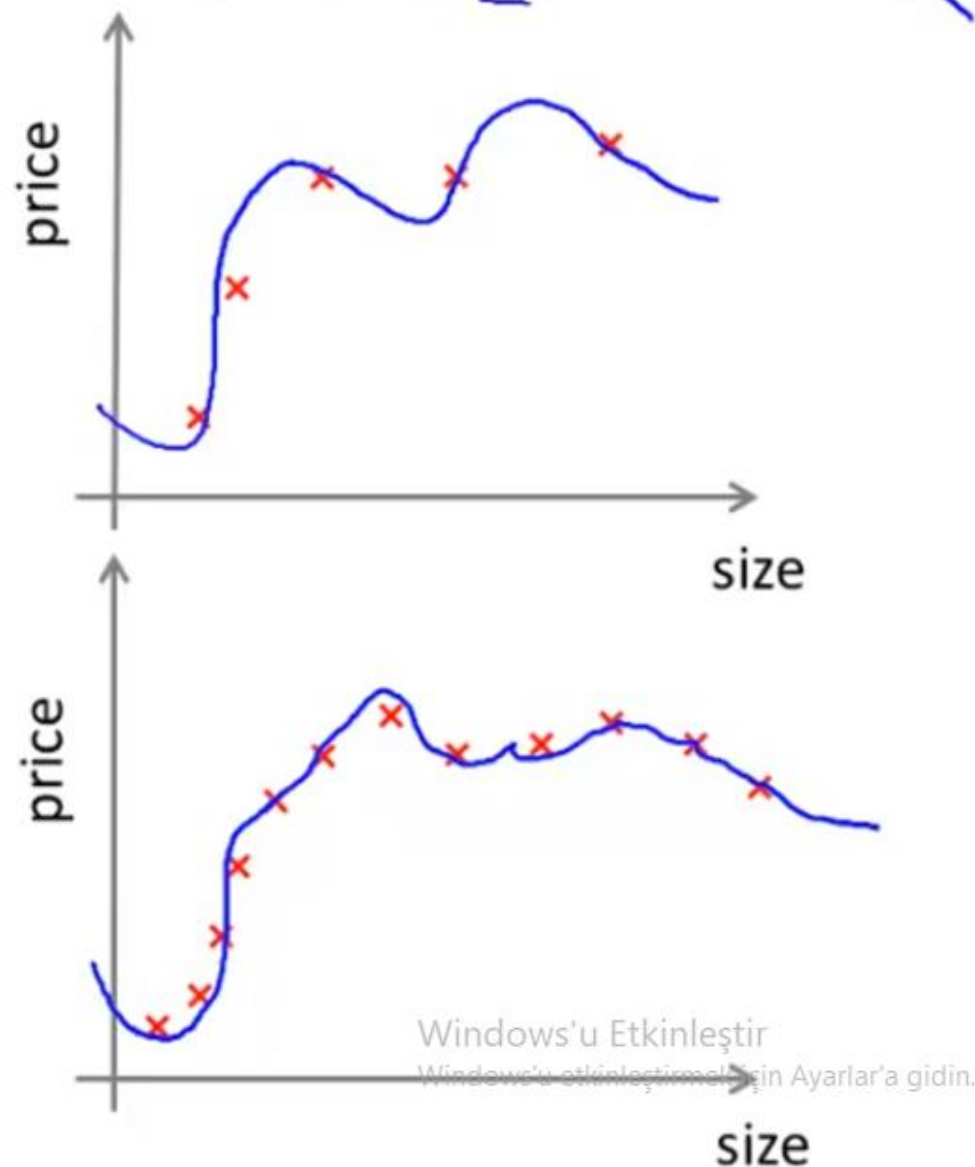
## High variance



If a learning algorithm is suffering from high variance, getting more training data is likely to help.

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(and small  $\lambda$ )



# Exercise

- In which of the following circumstances is getting more training data likely to significantly help a learning algorithm's performance?
  - ☐ Algorithm is suffering from high bias.
  - ☐ Algorithm is suffering from high variance.
  - ☐  $J_{CV}(\theta)$  (cross validation error) is much larger than  $J_{train}(\theta)$  (training error).
  - ☐  $J_{CV}(\theta)$  (cross validation error) is about the same as  $J_{train}(\theta)$  (training error).