Debugging a learning algorithm

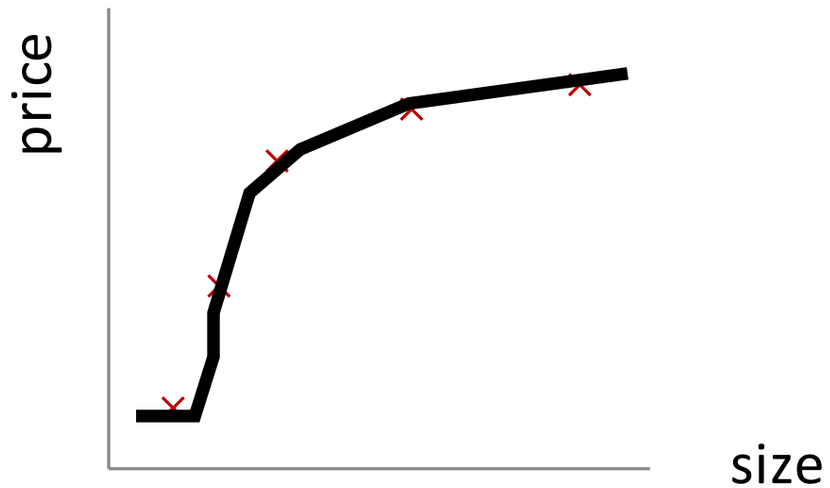
$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{m}\theta_j^2\right]$$

- Debugging a learning algorithm
- Get more training examples
- Try smaller sets of features
- Try getting additional features
- Try adding polynomial features
- Try decreasing/increasing regularization para.

Machine Learning Diagnostic

A test that you can run to gain insight what is/ isn't working with a learning algorithm, and gain guidance as to how best to improve its performance.

Evaluating your hypothesis



$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

Fails to generalize to new examples not in training set.

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

Training data

Test data

Compute test set error

Test set error in Linear Regression

Test set error in Logistic Regression

Model Selection

- Different model choices

1. $h_\theta(x) = \theta_0 + \theta_1 x$
2. $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$
3. $h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_3 x^3$

10. $h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{10} x^{10}$

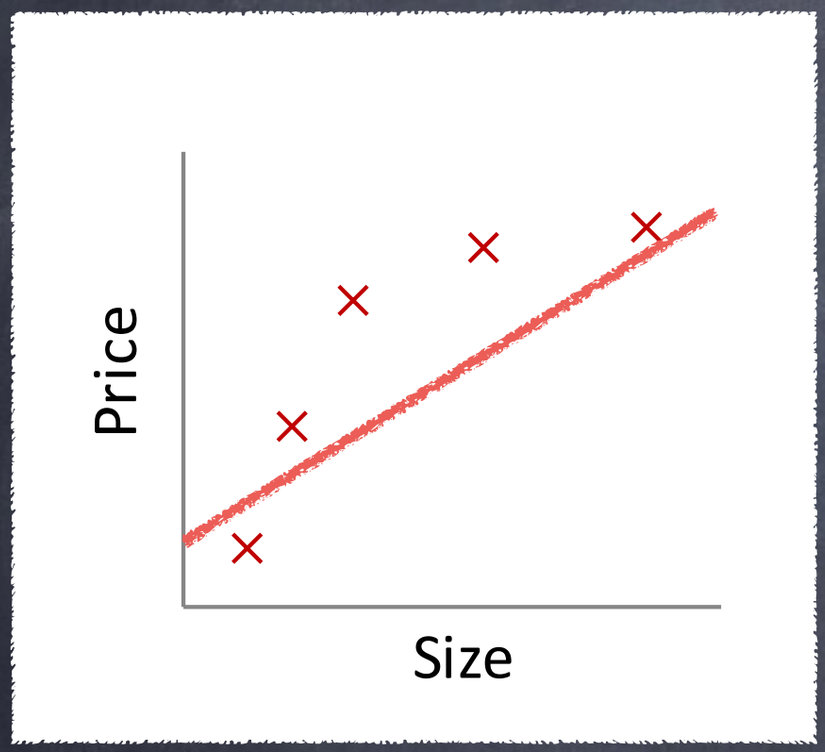
- Choose model with lower test er
- Problem: likely to be an optimist generalization error. I.e. our ext fit to test set.

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	
1985	300
1534	
1427	199
1380	
1494	243

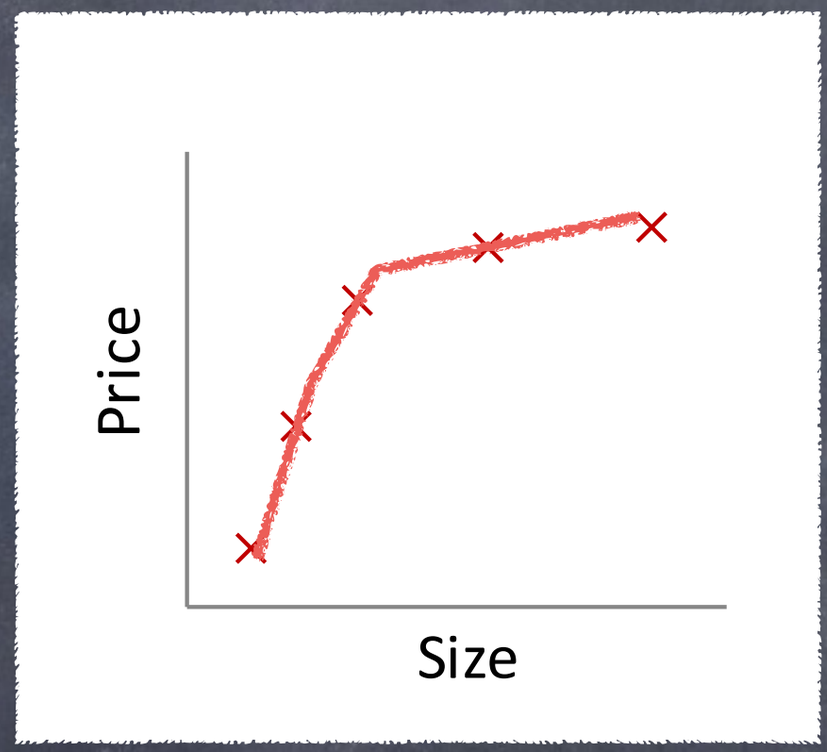
Training data

Cross-validation data

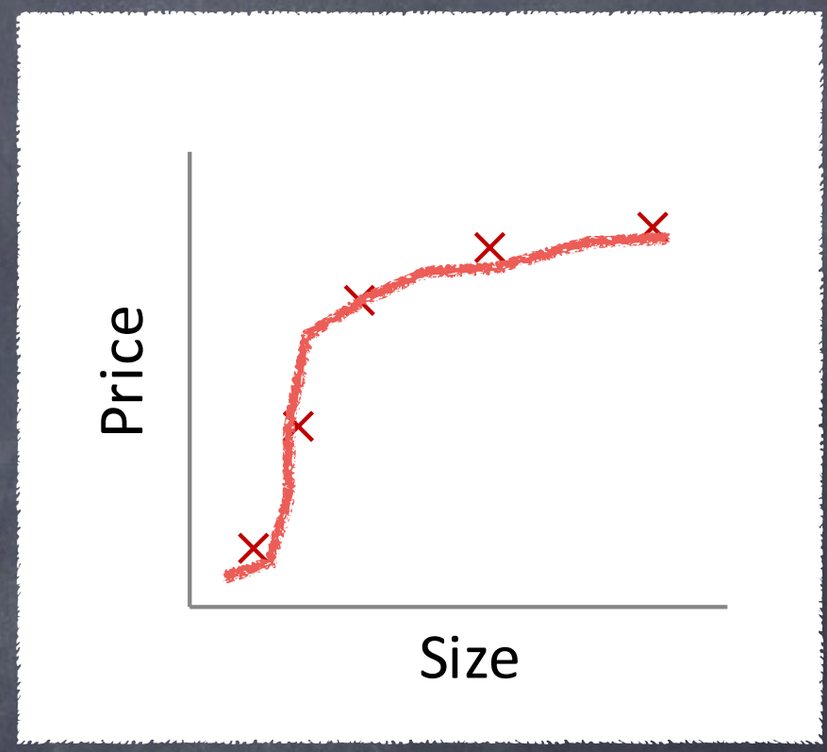
Testing data



Underfit

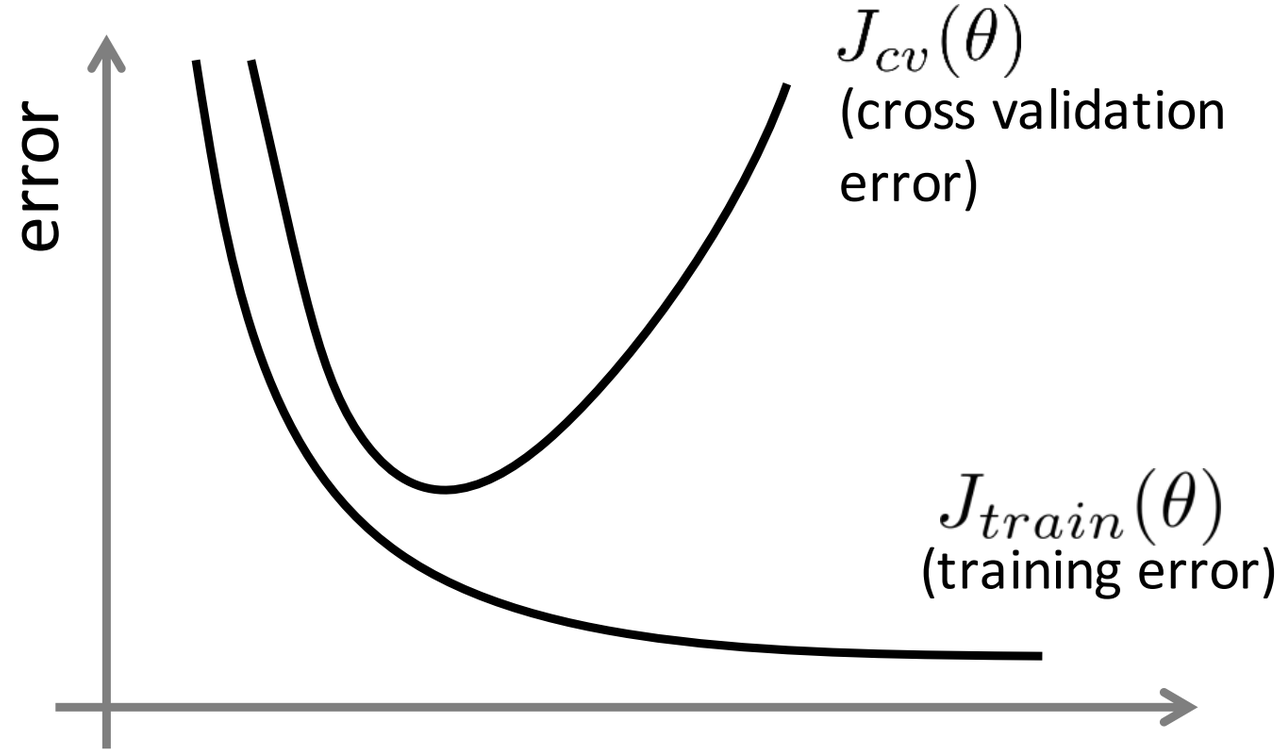


Just Right



Overfit

Bias vs. Variance



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

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

Linear Regression with Regularization

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^{m} \theta_j^2$$

1. Try $\lambda = 0$
2. Try $\lambda = 0.01$
3. Try $\lambda = 0.02$
4. Try $\lambda = 0.04$
5. Try $\lambda = 0.08$

12. Try $\lambda = 10$

bias vs. variance as a function of the regularization parameter?