

Evaluating a Learning Algorithm

- ✓ **Video:** Deciding What to Try Next
5 min
- ✓ **Video:** Evaluating a Hypothesis
7 min
- ✓ **Reading:** Evaluating a Hypothesis
4 min
- ✓ **Video:** Model Selection and Train/Validation/Test Sets
12 min
- ✓ **Reading:** Model Selection and Train/Validation/Test Sets
3 min

Bias vs. Variance

- ✓ **Video:** Diagnosing Bias vs. Variance
7 min
- ✓ **Reading:** Diagnosing Bias vs. Variance
3 min
- ✓ **Video:** Regularization and Bias/Variance
11 min
- ✓ **Reading:** Regularization and Bias/Variance
3 min
- ▶ **Video:** Learning Curves
11 min
- 📖 **Reading:** Learning Curves
3 min
- ▶ **Video:** Deciding What to Do Next Revisited
6 min
- 📖 **Reading:** Deciding What to do Next Revisited
3 min

Review

- ✓ **Reading:** Lecture Slides
10 min
- 📖 **Quiz:** Advice for Applying Machine Learning
5 questions
- 🔗 **Programming Assignment:** Regularized Linear Regression and Bias/Variance
3h

Building a Spam Classifier

- ▶ **Video:** Prioritizing What to Work On
9 min
- 📖 **Reading:** Prioritizing What to Work On
3 min

Regularization and Bias/Variance

Note: [The regularization term below and through out the video should be $\frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$ and **NOT** $\frac{\lambda}{2m} \sum_{j=1}^m \theta_j^2$]

Linear regression with regularization

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

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

Price vs Size plots illustrating bias and variance:

- Large λ → High bias (underfit)**
 $\lambda = 10000, \theta_1 \approx 0, \theta_2 \approx 0, \dots$
 $h_{\theta}(x) \approx \theta_0$
- Intermediate λ → "Just right"**
- Small λ → High variance (overfit)**
 $\lambda = 0$

Andrew Ng

In the figure above, we see that as λ increases, our fit becomes more rigid. On the other hand, as λ approaches 0, we tend to over overfit the data. So how do we choose our parameter λ to get it 'just right' ? In order to choose the model and the regularization term λ , we need to:

- Create a list of lambdas (i.e. $\lambda \in \{0, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24\}$);
- Create a set of models with different degrees or any other variants.
- Iterate through the λ s and for each λ go through all the models to learn some Θ .
- Compute the cross validation error using the learned Θ (computed with λ) on the $J_{CV}(\Theta)$ **without** regularization or $\lambda = 0$.
- Select the best combo that produces the lowest error on the cross validation set.
- Using the best combo Θ and λ , apply it on $J_{test}(\Theta)$ to see if it has a good generalization of the problem.

✓ Complete

Go to next item