

Classification and Representation

- ✓

Video: Classification

8 min
- ✓

Reading: Classification

2 min
- ✓

Video: Hypothesis Representation

7 min
- ✓

Reading: Hypothesis Representation

3 min
- ✓

Video: Decision Boundary

14 min
- ✓

Reading: Decision Boundary

3 min

Logistic Regression Model

- ✓

Video: Cost Function

10 min
- ✓

Reading: Cost Function

3 min
- ✓

Video: Simplified Cost Function and Gradient Descent

10 min
- ✓

Reading: Simplified Cost Function and Gradient Descent

3 min
- ✓

Video: Advanced Optimization

14 min
- ✓

Reading: Advanced Optimization

3 min

Multiclass Classification

- ✓

Video: Multiclass Classification: One-vs-all

6 min
- ✓

Reading: Multiclass Classification: One-vs-all

3 min

Review

- ✓

Reading: Lecture Slides

10 min
- ✓

Quiz: Logistic Regression

5 questions

Solving the Problem of Overfitting

- ✓

Video: The Problem of Overfitting

9 min
- ✓

Reading: The Problem of Overfitting

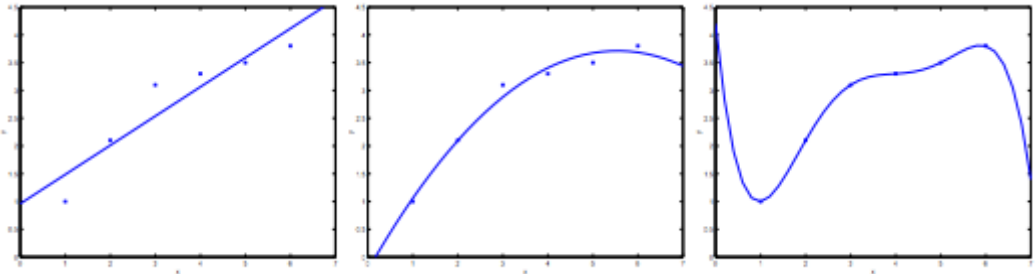
3 min

- ▶

Video: Cost Function

The Problem of Overfitting

Consider the problem of predicting y from $x \in \mathbb{R}$. The leftmost figure below shows the result of fitting a $y = \theta_0 + \theta_1 x$ to a dataset. We see that the data doesn't really lie on straight line, and so the fit is not very good.



Instead, if we had added an extra feature x^2 , and fit $y = \theta_0 + \theta_1 x + \theta_2 x^2$, then we obtain a slightly better fit to the data (See middle figure). Naively, it might seem that the more features we add, the better. However, there is also a danger in adding too many features: The rightmost figure is the result of fitting a 5th order polynomial $y = \sum_{j=0}^5 \theta_j x^j$. We see that even though the fitted curve passes through the data perfectly, we would not expect this to be a very good predictor of, say, housing prices (y) for different living areas (x). Without formally defining what these terms mean, we'll say the figure on the left shows an instance of **underfitting**—in which the data clearly shows structure not captured by the model—and the figure on the right is an example of **overfitting**.

Underfitting, or high bias, is when the form of our hypothesis function h maps poorly to the trend of the data. It is usually caused by a function that is too simple or uses too few features. At the other extreme, overfitting, or high variance, is caused by a hypothesis function that fits the available data but does not generalize well to predict new data. It is usually caused by a complicated function that creates a lot of unnecessary curves and angles unrelated to the data.

This terminology is applied to both linear and logistic regression. There are two main options to address the issue of overfitting:

- 1) Reduce the number of features:
- Manually select which features to keep.
- Use a model selection algorithm (studied later in the course).
- 2) Regularization
- Keep all the features, but reduce the magnitude of parameters θ_j .
- Regularization works well when we have a lot of slightly useful features.

✓ Complete

Go to next item

