Q

coursera

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

Review

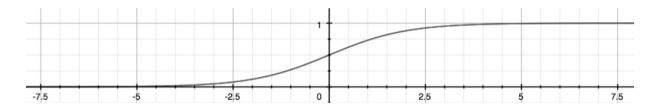
Hypothesis Representation

We could approach the classification problem ignoring the fact that y is discrete-valued, and use our old linear regression algorithm to try to predict y given x. However, it is easy to construct examples where this method performs very poorly. Intuitively, it also doesn't make sense for $h_{\theta}(x)$ to take values larger than 1 or smaller than 0 when we know that $y \in \{0, 1\}$. To fix this, let's change the form for our hypotheses $h_{\theta}(x)$ to satisfy $0 \le h_{\theta}(x) \le 1$. This is accomplished by plugging $\theta^T x$ into the Logistic Function.

Our new form uses the "Sigmoid Function," also called the "Logistic Function":

$$egin{aligned} h_{ heta}(x) &= g(heta^T x) \ z &= heta^T x \ g(z) &= rac{1}{1+e^{-z}} \end{aligned}$$

The following image shows us what the sigmoid function looks like:



The function g(z), shown here, maps any real number to the (0, 1) interval, making it useful for transforming an arbitrary-valued function into a function better suited for classification.

 $h_{\theta}(x)$ will give us the **probability** that our output is 1. For example, $h_{\theta}(x)=0.7$ gives us a probability of 70% that our output is 1. Our probability that our prediction is 0 is just the complement of our probability that it is 1 (e.g. if probability that it is 1 is 70%, then the probability that it is 0 is 30%).

$$egin{aligned} h_{ heta}(x) &= P(y=1|x; heta) = 1 - P(y=0|x; heta) \ P(y=0|x; heta) + P(y=1|x; heta) = 1 \end{aligned}$$

✓ Complete

Go to next item



