

**Classification and Representation****Logistic Regression Model****Multiclass Classification****Review****Solving the Problem of Overfitting**

- ✓ **Video:** The Problem of Overfitting  
9 min
- ✓ **Reading:** The Problem of Overfitting  
3 min
- ✓ **Video:** Cost Function  
10 min
- ✓ **Reading:** Cost Function  
3 min
- ✓ **Video:** Regularized Linear Regression  
10 min
- ✓ **Reading:** Regularized Linear Regression  
3 min
- ▶ **Video:** Regularized Logistic Regression  
8 min
- 📖 **Reading:** Regularized Logistic Regression  
3 min
- Review**
- ✓ **Reading:** Lecture Slides  
10 min
- 📋 **Quiz:** Regularization  
5 questions
- 📁 **Programming Assignment:** Logistic Regression  
3h



# Regularized Linear Regression

**Note:** [8:43 - It is said that  $X$  is non-invertible if  $m \leq n$ . The correct statement should be that  $X$  is non-invertible if  $m < n$ , and may be non-invertible if  $m = n$ .

We can apply regularization to both linear regression and logistic regression. We will approach linear regression first.

## Gradient Descent

We will modify our gradient descent function to separate out  $\theta_0$  from the rest of the parameters because we do not want to penalize  $\theta_0$ .

$$\begin{aligned} &\text{Repeat } \{ \\ &\quad \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_0^{(i)} \\ &\quad \theta_j := \theta_j - \alpha \left[ \left( \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right] \quad j \in \{1, 2, \dots, n\} \\ &\} \end{aligned}$$

The term  $\frac{\lambda}{m} \theta_j$  performs our regularization. With some manipulation our update rule can also be represented as:

$$\theta_j := \theta_j \left( 1 - \alpha \frac{\lambda}{m} \right) - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)}$$

The first term in the above equation,  $1 - \alpha \frac{\lambda}{m}$  will always be less than 1. Intuitively you can see it as reducing the value of  $\theta_j$  by some amount on every update. Notice that the second term is now exactly the same as it was before.

## Normal Equation

Now let's approach regularization using the alternate method of the non-iterative normal equation.

To add in regularization, the equation is the same as our original, except that we add another term inside the parentheses:

$$\theta = \left( X^T X + \lambda \cdot L \right)^{-1} X^T y$$

where  $L = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \\ & & & & 1 \end{bmatrix}$

$L$  is a matrix with 0 at the top left and 1's down the diagonal, with 0's everywhere else. It should have dimension  $(n+1) \times (n+1)$ . Intuitively, this is the identity matrix (though we are not including  $x_0$ ), multiplied with a single real number  $\lambda$ .

Recall that if  $m < n$ , then  $X^T X$  is non-invertible. However, when we add the term  $\lambda \cdot L$ , then  $X^T X + \lambda \cdot L$  becomes invertible.

✓ Complete

Go to next item

