

# 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  $\theta_0$ .

Repeat {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m} \theta_j \quad j \in \{1, 2, \dots, n\}$$

}

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

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

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:

