**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**

**Solving the Problem of Overfitting**

- **Video:** The Problem of Overfitting
  9 min
- **Reading:** The Problem of

# 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 $\theta_0$.

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

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

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

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:

$$
\begin{aligned}
&\theta = \left(X^TX + \lambda \cdot L\right)^{-1} X^Ty \\
&\text{where } L = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}
\end{aligned}
$$

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)×(n+1). Intuitively, this is the identity matrix (though we are not including $x_0$), multiplied with a single real number λ.

Recall that if m < n, then $X^TX$ is non-invertible. However, when we add the term λ⋅L, then $X^TX$ + λ⋅L becomes invertible.

✓ Complete

Go to next item