1. Randomly initialize weights

2. Implement forward propagation to get $h_\Theta(x^{(i)})$ for any $x^{(i)}$

3. Implement code to compute cost function $J(\Theta)$

4. Implement backprop to compute partial derivatives $\frac{\partial}{\partial \Theta_{jk}^{(l)}} J(\Theta)$

5. Use gradient checking to compare $\frac{\partial J(\Theta)}{\partial \Theta_{jk}^{(l)}}$ computed using backpropagation vs. using numerical estimate of gradient of $J(\Theta)$

6. Use gradient descent or advanced optimization method with backpropagation to try to minimize $J(\Theta)$ as a function of parameters Θ