\[ x = \text{model}(x) \]

\[ z_1 = x W^1 \text{ activating } + b \]

\[ A_1 = \max(z_1, 0) \]

\[ z_2 = A_1 W^2 \text{ activating } + b \]

\[ A_2 = \max(z_2, 0) \]

\[ z_{10} = A_2 W^3 \text{ activating } + b \]

\[ L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2 \]

\[ J(\hat{y}, y) = L(\hat{y}, y) + \lambda \| \Theta \|_2 \]

\[ \frac{\partial J}{\partial \Theta} = \frac{\partial L}{\partial \Theta} + \lambda \Theta \]