Runqiu Ye
Stanford CS299
Problem Set #3
07/29/2024

Problem Set #3: Deep Learning & Unsupervised Learning

Problem 1 A simple neural network
Let $X = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ be dataset of m examples with 2 features. That is, $x^{(i)} \in \mathbb{R}^2$. Samples are classifed into 2 categorie with labels $y \in \{0, 1\}$, as shown in Figure 1. Want to perform binary classification using a simple neural networks with the architecture shown in Figure 2.
Two features x_1 and x_2, the three neurons in the hidden layer h_1, h_2, h_3, and the output neuron as o. Weight from x_i to h_j be $w_{i,j}^{[1]}$ for $i = 1, 2$ and $j = 1, 2, 3$, and weight from h_j to o be $w_j^{[2]}$. Finally, denote intercept weight for h_j as $w_{0,j}^{[1]}$ and the intercept weight for o as $w_0^{[2]}$. Use average squared loss instead of the usual negative log-likelihood:

$$l = \frac{1}{m} \sum_{i=1}^{m} (o^{(i)} - y^{(i)})^2.$$

(a) Suppose we use sigmoid function as activation function for h_1, h_2, h_3, and o. We have

$$h_1 = g(w_1^{[1]} x), \quad h_2 = g(w_2^{[1]} x), \quad h_3 = g(w_3^{[1]} x), \quad o = g(w_2^{[2]} h).$$

Hence,

$$\frac{\partial l}{\partial w_{1,2}^{[1]}} = \frac{1}{m} \sum_{i=1}^{m} 2(o - y) o (1 - o) w_2^{[2]} h_2 (1 - h_2) x_1,$$

where $h_2 = g(w_{1,2}^{[1]} x_1 + w_{2,2}^{[1]} x_2 + w_{3,2}^{[1]} x_3)$ and g is the sigmoid function. Therefore, the gradient descent update to $w_{1,2}^{[1]}$, assuming learning rate α is

$$w_{1,2}^{[1]} := w_{1,2}^{[1]} - \frac{2\alpha}{m} \sum_{i=1}^{m} (o - y) o (1 - o) w_2^{[2]} h_2 (1 - h_2) x_1,$$

where $h_2 = g(w_{1,2}^{[1]} x_1 + w_{2,2}^{[1]} x_2 + w_{3,2}^{[1]} x_3)$.

■