CS229 Fall 2017
Problem Set #3 Solutions: Deep Learning & Unsupervised Learning

Author: LFhase rimemosa@163.com

A Simple Neural Network

(a) Using Chain Rule, we know that

$$\frac{\partial loss}{\partial w_{1,2}^{[1]}} = \frac{\partial loss}{\partial o}\frac{\partial o}{\partial h_2}\frac{\partial h_2}{\partial w_{1,2}^{[1]}}$$

let $g(x)$ denote the sigmoid function, then we have

$$g'(x) = g(x)(1 - g(x))$$

so

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

where

$$h_2^{(i)} = g(x_1^{(i)}w_{1,2}^{[1]} + x_2^{(i)}w_{2,2}^{[1]} + w_{0,2}^{[1]})$$

(b) let $(0.5, 0.5)$, $(3.5, 0.5)$, $(0.5, 3.5)$ be the three poinst of the triangle. The forward transport in the neural network can be written in matrix form.

$$\begin{bmatrix} -1.5 & 3 & 0 \\ -1.5 & 0 & 3 \\ 9 & -3 & 3 \end{bmatrix} \times \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

and

$$\begin{bmatrix} -1 & -1 & -1 & 2.33 \end{bmatrix} \times \begin{bmatrix} 1 \\ h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

Once the point is in the triangle, the first product will be

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

So the second product will be -0.67 and the final result will be 0. Otherwise, the second product will be larger or equal to 0.33 and the final result will be 1.

(c) Using $f(x) = x$ as hidden layer activation function, we can see the neural network as **a simple neural network without hidden layer**, who only has the **convex boundary** and can't deal with the problem described in statement.

EM for MAP estimation

(a) Using Chain Rule, we know