

IOWA STATE UNIVERSITY

DEPARTMENT OF ELECTRICAL AND COMPUTER
ENGINEERING

DEEP MACHINE LEARNING: THEORY AND PRACTICE

EE 526X

Homework 2

Author:
Vishal DEEP

Instructor:
Dr. Zhengdao WANG

October 5, 2019



1 Problem 1

We have a softplus function,

$$f(z) = \log(1 + e^z) \quad (1)$$

Taking first derivative of the function:

$$f'(z) = \frac{1}{(1 + e^z)} e^z \quad (2)$$

multiplying and dividing by e^{-z} ,

$$f'(z) = \frac{e^z \cdot e^{-z}}{(1 \cdot e^{-z} + e^z \cdot e^{-z})} = \frac{1}{(1 + e^{-z})} \quad (3)$$

Now taking second derivative of softplus function,

$$f''(z) = \frac{(1 + e^{-z}) \cdot 0 + e^{-z}}{(1 + e^{-z})^2} \quad (4)$$

$$f''(z) = \frac{e^{-z}}{(1 + e^{-z})^2} \quad (5)$$

From eq. (5), numerator is always positive because exponential is a positive number and any power to a positive is number is always a positive number. The denominator is always positive because is squared. Therefore we can say that second derivative of a softplus function is always positive for every value of z . This implies that a softplus function is convex in z .

2 Problem 2

We have a vector $z = [z_1, z_2, \dots, z_n]^T$ and $p = [p_1, p_2, \dots, p_n]^T$, where p is output after z is applied to a softmax function.

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (6)$$

The jacobian matrix will have two types of elements diagonal and off-diagonal, we will calculate both and then generalize the result for all elements.

Diagonal row elements:

$$\begin{aligned} \frac{\partial p_i}{\partial z_i} &= \frac{\sum_{j=1}^n e^{z_j} \cdot e^{z_i} - e^{z_i} \cdot e^{z_i}}{(\sum_{j=1}^n e^{z_j})^2} \\ \frac{\partial p_i}{\partial z_i} &= \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} - \left(\frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \right)^2 \\ \frac{\partial p_i}{\partial z_i} &= p_i - p_i^2 \end{aligned} \quad (7)$$

off-diagonal row elements:

$$\begin{aligned}\frac{\partial p_i}{\partial z_j} &= \frac{\sum_{j=1}^n e^{z_j} \cdot 0 - e^{z_i} \cdot e^{z_j}}{(\sum_{j=1}^n e^{z_j})^2} \\ \frac{\partial p_i}{\partial z_j} &= -\frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \times \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \\ \frac{\partial p_i}{\partial z_j} &= -p_i p_j\end{aligned}\tag{8}$$

The Jacobian matrix is given by

$$\frac{\partial p}{\partial z} = \begin{bmatrix} \frac{\partial p_0}{\partial z_0} & \frac{\partial p_0}{\partial z_1} & \cdots & \frac{\partial p_0}{\partial z_n} \\ \frac{\partial p_1}{\partial z_0} & \frac{\partial p_1}{\partial z_1} & \cdots & \frac{\partial p_1}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_n}{\partial z_0} & \frac{\partial p_n}{\partial z_1} & \cdots & \frac{\partial p_n}{\partial z_n} \end{bmatrix}\tag{9}$$

Using eq. (7) and eq. (8), we get

$$\frac{\partial p}{\partial z} = \begin{bmatrix} p_0 - p_0^2 & -p_0 p_1 & \cdots & -p_0 p_n \\ -p_1 p_0 & p_1 - p_1^2 & \cdots & -p_1 p_n \\ \vdots & \vdots & \ddots & \vdots \\ -p_n p_0 & -p_n p_1 & \cdots & p_n - p_n^2 \end{bmatrix}\tag{10}$$