CSC411: Assignment #$\sqrt{1}$

Due on Sunday, January 29, 2017

Naireen Hussain

February 8, 2018

Problem 1

Part 1: Describing the Data

Problem 2

Single Layer Neural Net

Problem 3

Deriving the Gradient Descent
The derivative of the cost function with respect to each weight W_{ij} is shown as follows
For a single layer, the derivative has the formula:

$$\frac{\partial C}{\partial W^{i,j}} = \frac{\partial g(z)}{\partial W^{i,j}} \frac{\partial C}{\partial g} \textbf{ where } g(z) = g(\sum_j W^{i,j} x_j + b^j) \tag{1}$$

$$\frac{\partial C}{\partial g(z)} \frac{\partial g(z)}{\partial z_i} \frac{\partial z_i}{\partial W^{i,j}} \tag{2}$$

To show the first term:
Since the cost function is the log likelihood function, which is defined as $-\Sigma_j y_j log p_j$, where p_j are the probabilities per class determined by a soft max function.

$$p_i = \frac{e^{g_i}}{\Sigma_j e^{g_j}} \tag{3}$$

$$\frac{\partial p_i}{\partial g_i} = p_i(1 - p_i) \tag{4}$$

$$C = -\Sigma_j y_j log p_j \tag{5}$$

$$\frac{\partial C}{\partial p_j} = -\frac{y_j}{p_j} \tag{6}$$

$$\frac{\partial C}{\partial g_i} = \frac{\partial C}{\partial p_i} \frac{\partial p_i}{\partial g_i} = -\frac{y_i}{p_i} p_i(1 - p_i) = -y_i(1 - p_i) \tag{7}$$

For the second term, since the activation function is identity, $\frac{\partial g}{\partial z} = 1$
For the third term, this is simply the value x_j
Combining the three terms yield:

$$\frac{\partial C}{\partial g(z)} \frac{\partial g(z)}{\partial z} \frac{\partial z_i}{\partial W^{i,j}} = -y_j(1 - p_j)(x_j) \tag{8}$$