Deep Learning Homework 1 Math Questions

October 2019

1 Hinge loss

Suppose you have a binary classification problem with training data: $\{x^i, y^i\}$, and you want to use a linear hypothesis

$h_\theta(x)$ with nonlinear features of the input in a sixth order polynomial. You decide to use hinge loss, which is defined for each data sample in your training set as follows:

$l(\theta) = max(1 - yh_\theta(x), 0)$. Devise an optimization based on Gradient Descent to solve this problem. Clearly state the hypothesis function, the derivation for the loss function, and the pseudo-algorithm for the optimization.

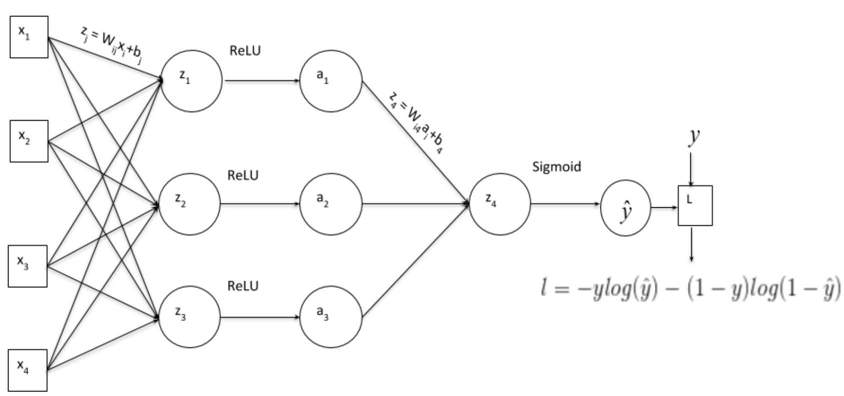

Figure 1: Fully Connected Neural Network Architecture

2 Derive the gradients

You build a Fully Connected Neural Network components of which are as shown in the Figure 1.

2.1

Derive the gradient of the loss L with respect to (wrt) the output of second layer: z_4.

2.2

Given upstream gradient $\delta z_4 = \frac{\partial L}{\partial z_4}$, derive gradient of the loss wrt the second layer weight W_{14} of layer: $\frac{\partial L}{\partial W_{14}}$

2.3

Given upstream gradient $\delta a_1 = \frac{\partial L}{\partial a_1}$, derive gradient of the loss wrt the first layer weight W_{11} of layer: $\frac{\partial L}{\partial W_{11}}$

2.4

How would gradients change in first three questions, if we add L2 Weight Regularization? Write three gradients again separately.