Deep Learning Assignment 1

Yun-shao Sung* and Chung-Ling Yao†

Abstract. This is the report for deep learning assignment 1

1. Warmup. Write $\frac{\partial E}{\partial X_{in}}$ in terms of $\frac{\partial E}{\partial X_{out}}$

$$\frac{\partial E}{\partial X_{in}} = \frac{\partial E}{\partial X_{out}} \frac{\partial F(X_{in}, W_i)}{\partial X_{in}} = \frac{\partial E}{\partial X_{out}} \frac{e^{X_{in}}}{(1+e^{X_{in}})^2} = \frac{\partial E}{\partial X_{out}} X_{out}(1 - X_{out}) \tag{1.1}$$

2. Multinomial logistic regression. Write the expression of $\frac{\partial (X_{out})_i}{\partial (X_{in})_j}$ if $i = j$, , and let $C = \sum_k e^{(X_I)_k} - e^{(X_I)_i}$

$$(X_o)_i = \frac{e^{(X_I)_i}}{\sum_k e^{(X_I)_k}} = \frac{e^{(X_I)_i}}{e^{(X_I)_0} + e^{(X_I)_i} + \ldots e^{(X_I)_i} + \ldots + e^{(X_I)_k}} = \frac{e^{(X_I)_i}}{C + e^{(X_I)_i}} \tag{2.1}$$

$$\frac{\partial (X_o)_i}{\partial (X_I)_i} = \frac{\partial}{\partial (X_I)_i} \left(\frac{e^{(X_I)_i}}{C + e^{(X_I)_i}} \right) = \frac{-\beta e^{-\beta (X_I)_i}}{C + e^{(X_I)_i}} + \frac{\beta e^{-2\beta (X_I)_i}}{(C + e^{(X_I)_i})^2} = \beta X_o(-1 + X_o) \tag{2.2}$$

if $i \neq j$, and let $K = \sum_k e^{(X_I)_k} - e^{(X_I)_j}$

$$\frac{\partial (X_o)_i}{\partial (X_I)_j} = \frac{\partial}{\partial (X_I)_j} \left(\frac{e^{(X_I)_i}}{K + e^{(X_I)_j}} \right) = \frac{\beta e^{-\beta (X_I)_i} e^{-\beta (X_I)_j}}{(K + e^{(X_I)_i})^2} = \beta (X_o)_i (X_o)_j \tag{2.3}$$

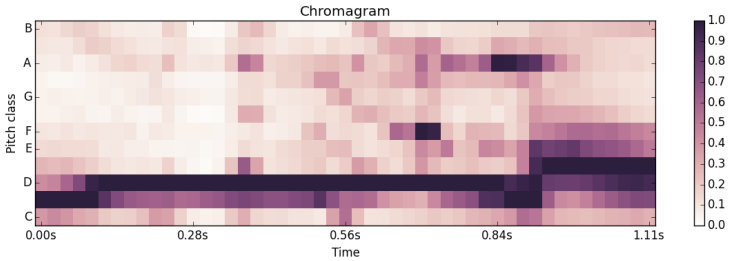

Figure 2.1. *Histogram of accuracy result from search space. Top row is baseline method 5, and the remaining rows are new feature extraction method 1, 2, and 3 respectively*

*yss265@nyu.edu
†cly264@nyu.edu