CMPT 728 Deep Learning

School of Computing Science

Simon Fraser University

Instructor : Oliver Schulte

Assignment 1:

This assignment consists of two parts : conceptual exercises + a project.

You should submit a assignment1.zip package including :
- Answers to the conceptual exercises part in .PDF version ;
- Your code of project part in both .PDF and .ipynb (Jupyter notebook) version.

A. Conceptual Exercises (50 points)

A.1 Gradients and Backpropagation (5 +5 = 10 points)

Given a training example *(**x**,y)*, and a set of linear weights **w**, find the gradient of the loss function with respect to **w** for the following loss functions (definition in the slides).

1. Least-squares error (see also Exercise 1.6 in the text)
2. Cross-entropy

Describe informally but clearly how you can incorporate these gradients in the backpropagation algorithm described in class (see lecture slides).

A.2 Broadcasting (10 points)

Compute the following:

$$\begin{pmatrix} 0.5 & 1.5 \\ 3.5 & 2 \end{pmatrix} \begin{pmatrix} 0.5 & 1 \\ 3 & 2 \end{pmatrix} + \begin{pmatrix} 1.25 & 3 \end{pmatrix}$$

You should assume broadcasting so the computation is well defined.

A.3 Trace Backpropagation (30 points)

Consider a neural net with one hidden layer, two inputs *a* and *b*, one hidden unit *c*, and one output unit *d*. The activation function is the sigmoid for each node. This network has five weights (w_{ac}, w_{bc}, w_{0c}, w_{cd}, w_{0d}), where w_{0x} represents the bias or threshold weight for unit *x*. Initialize these weights to the values (.2, .1, .2, .1,.2), then give their values after each of the first two training iterations of Backpropagation algorithm. Assume learning rate (step size) of 0.1, stochastic (incremental) gradient descent (without momentum), least-squares for the loss/error function. The input are the following training examples:

Data Point	a	b	d
x_1	1	0	1
x_2	0	1	0

1. Using the notation in the slides, show the formulas for computing the following quantities, for each node *x:*
   - a_x
   - $\Delta[x]$
   - Weight update for w_{xy}

2. Fill in the following table using the formulas from the slides. You can expand this to include more information (e.g. derivatives of activation functions) if you like.

Data Point	a_c	$\Delta[c]$	a_d	$\Delta[d]$	w_{0c}	w_{ac}	w_{bc}	w_{cd}	w_{0d}
x_1									
x_2									

B. Project Part (100 points):

Please see the assignment1_proj.ipynb and the assignment1_proj.pdf files.