University of Toronto Mississauga
Department of Mathematical and Computational Sciences
CSC 321 — Introduction to Neural Networks and Machine Learning, Spring 2014

Midterm Test

Name:

Student Number:

This test is closed book, but you are allowed one page of notes (single-sided, 8.5x11 inch) in 12-point font (or larger) and no more than 6000 characters. No other aids are allowed.

If you leave a question blank and simply write "I don't know," you will receive 20% of the value of the question.

Write all answers on the test booklet, using the back of pages if necessary. 50 minutes. 10 pages. 4 questions. 65 points. Good Luck!

1. (10 points total) **Batch and on-line learning**.

   In the following, assume a neural network with a single output using squared error. Be sure to define all your symbols.

   (a) (2 points) Write down the error function for online learning.

   (b) (2 points) Write down the error function for batch learning.

(c) (3 points) Briefly describe on-line learning.

(d) (3 points) Briefly describe batch learning.

2. (25 points total) **Overfitting**.

   (a) (5 points) Briefly describe overfitting and why it is a problem.

(b) (5 points) Draw a simple example showing well-fitted and overfitted data and label the various parts.

(c) (15 points) Briefly describe three ways of preventing overfitting in neural networks and why they work.

This page is for answers and rough work

3. (15 points) **Back propagation**.

Consider a 1-layer neural net with three input units, 1 output unit, no hidden units and no bias terms. Suppose that the output unit uses a sigmoid activation function, *i.e.*, $y = 1/(1 + e^{-z})$, where z is the total input to the unit. Let y be the computed output of the neural net, let d be the desired output, and let $C = -d \log y - (1 - d) \log (1 - y)$ be the cross entropy error. Write down the equations for a single step of weight updates by gradient descent (based on a single data sample), and derive all the necessary derivatives. Simplify your answers, and be sure to clearly identify all the variables you use.

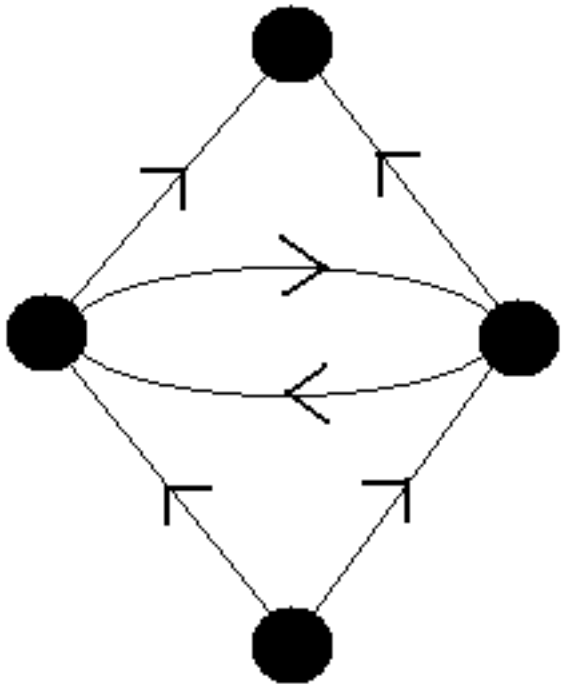
Hint: use the chain rule and recall the following results:

$$\frac{\partial y}{\partial z} = y(1 - y) \qquad\qquad \frac{\partial \log u}{\partial u} = \frac{1}{u}$$

This page is for answers and rough work

4. (15 points total) **Recurrent neural networks**.

   Consider the following recurrent neural network, in which the bottom node is the input, the top node is the output, and the middle two nodes are hidden.



   (a) (6 points) Draw an equivalent, feed-forward (non-recurrent) neural network representing three time instances.

   (b) (6 points) Add weights to your feed-forward network, clearly indicating which weights are equal.

   (c) (3 points) Mark the corresponding weights on the recurrent network above.

This page is for answers and rough work