Name ____________________________________

CSE 5526 - Autumn 2014

Midterm

Directions:

1. This is a closed-book, closed-note test, except for one 8.5x11” sheet of notes.
2. You may not consult with any other person.
3. You may not use any communication device or computer
4. You have 80 minutes to finish.
5. Write all work on the test paper. Use reverse side if needed (clearly indicate so).
6. There are *three* problems, with a total of 100 points, *plus* a bonus problem (10 points)

Problem 1. Short Answers (40 points)

(a) (10 points) Explain why backpropagation cannot be used for learning in binary hidden units

(b) (10 points) What is the advantage of using cross-validation over splitting a dataset into dedicated training and test sets? When is that less important?

(c) (10 points) Describe 3 optimization tricks for speeding up learning in multi-layer perceptron training for a fixed error function and network design.

(d) (10 points) Compute the gradient of $f(x, y) = 3x^2 + 2xy - y^2$ at the point $(2, 2)^T$.

Problem 2. (30 points) Consider the following two sets of points: $C_1 = \{(0,0), (-1,1), (1,1)\}$, $C_2 = \{(0,2), (-2,0), (2,0)\}$.

a) Are these points linearly separable? Why or why not?

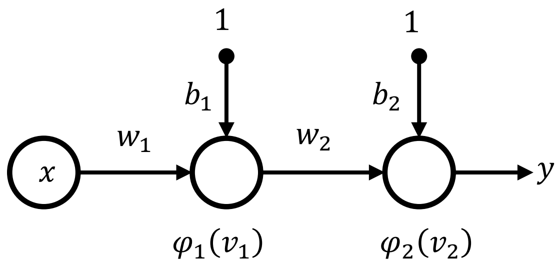
b) Design a network of McCulloch-Pitts neurons that can separate them and plot their decision boundaries.

c) Apply the transformation $\varphi(x) = \sum_i x_i^2$ to each of the points, where x_i is the ith dimension of point x.

d) Are the transformed points linearly separable? Why or why not?

e) Design a network of McCulloch-Pitts neurons that can separate them.

Problem 3. (30 points) Consider the following two-layer MLP with a *single* neuron in each layer where $w_1 = 1$, $b_1 = -2$, $w_2 = 2\ln 3$, $b_2 = 0$, and $\varphi_1(x) = \varphi_2(x) = \sigma(x) = \frac{1}{1+e^{-x}}$



a) Write down the equation for computing the network output, y, in terms of the parameters of the network and the input, x. You do not need to expand the $\varphi()$ functions.

b) For the input $x = 2$, compute the output of the network (the various parameters have been chosen so that you should be able to do it without a calculator).

c) Write down the backpropagation equations for updating the parameters of the network, w_2, b_2, w_1, b_1 based on the sum of square errors criterion for a general input x, desired output d, and learning rate η

d) Compute the updates for these parameters for the input $x = 2$ with desired output $d = 1$, and learning rate $\eta = 1$.

Bonus Problem. (10 points) Say that we want to train a two-layer MLP, but in addition to the usual sum of square errors cost function, we also add in a penalty term on large weights

$$E = \frac{1}{2}\left(\sum_k (d_k - y_k)^2\right) + \frac{\alpha}{2}\sum_{ij} w_{ij}^2 + \frac{\alpha}{2}\sum_{jk} w_{jk}^2$$

Where α is a user-defined constant. The penalty on large weights is thus of the form of the sum over all the weights in the network of each weight squared. This is a regularization technique known as "weight decay".

a) Write down the backpropagation update rules to train this network to minimize this new cost function.

b) What effect will increasing the α parameter have on the weights that are learned?