Name ____________________________________

CSE 5526 - Autumn 2014

Practice Midterm

Directions for the real midterm on 10/14:

1. The real midterm will be a closed-book, closed-note test, except for a single 8.5x11" sheet of notes that you are allowed to bring.
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) Describe the perceptron learning rule verbally (i.e. not quantitatively).

(b) (10 points) Give an example learning machine that achieves zero variance.

(c) (10 points) Based on backpropagation learning for multilayer perceptrons, give *two* reasons why simple (single-layer) perceptron learning does not generalize to multiple layers in a straightforward manner (i.e. why backpropagation is considered a breakthrough).

(d) (10 points) Compute the gradient of $f(x, y) = 2x + y^3$ at the point $(0, 1)^T$.

Problem 2. (30 points) In a multi-class classification problem, if N datapoints in K classes can be separated by $K(K–1)/2$ (hyper)planes, each of which separates a pair of classes, then they are *pairwise linearly separable*. Let the discriminant functions for a 3-class classifier be

$$g_{12}(\mathbf{x}) = -x_1 - x_2 + 5$$
$$g_{13}(\mathbf{x}) = -x_1 + 3$$
$$g_{23}(\mathbf{x}) = -x_1 + x_2$$

and $g_{ji}(\mathbf{x}) = -g_{ij}(\mathbf{x})$. Here there are two input features, x_1 and x_2. The decision rule is: assign $\mathbf{x}$ to C_k if and only if $g_{kj}(\mathbf{x}) > 0$ for all $j \neq k$.

Draw the decision boundaries and label classified regions and any unclassified regions. Classify the two points $\mathbf{x} = (4, 3)^T$ and $\mathbf{x} = (0, 2)^T$. If there is an unclassified region, prove it by finding a point that does not get classified according to the above rule.

Problem 3. (30 points) Show that the function $\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$ and the logistic sigmoid function $\sigma(a) = \frac{1}{1+e^{-a}}$ are related by

$$\tanh(a) = 2\sigma(2a) - 1.$$

Hence show that a general linear combination of logistic sigmoid functions of the form

$$y(x, \boldsymbol{w}) = w_0 + \sum_j w_j \sigma\left(\frac{x - \mu_j}{s}\right)$$

is equivalent to a linear combination of tanh functions of the form

$$y(x, \boldsymbol{u}) = u_0 + \sum_j u_j \tanh\left(\frac{x - \mu_j}{2s}\right)$$

and find expressions to relate the new parameters $\{u_0, \dots, u_M\}$ to the original parameters $\{w_0, \dots, w_M\}$.

Bonus Problem. (10 points) Given the sum of squared errors for the cost function E:

$$E = \frac{1}{2}\sum_i (d_i - y_i)^2$$

where i indexes an input pattern. Derive the weight update rule based on gradient descent. In the above equation, $y_i = \varphi(v_i)$, and y_i is the output of the sole output unit (see the following figure). v_i is the activation potential to the output unit, defined as $v_i = \boldsymbol{w}^T \boldsymbol{x}_i$, and $\boldsymbol{x}_i$ is an input vector. The activation function $\varphi()$ is assumed to be differentiable, and d_i is the desired output of the unit for the input pattern $\boldsymbol{x}_i$.