ECS 271 Machine Learning, Spring 2004

Homework #3
Due 22 April 2004

(1) (10 points) On a separate sheet (separate from the rest of the answers), please give an updated term paper proposal. This is the last chance for you to make a change of the topic, if you need to. What I want is a concise statement of the problem (what is it you want to investigate), identification of a potential method(s), and some insight into the data set you will be using. As ML is primarily a method to learn from data, it is absolutely essential you know what the data set is by now. Then, I also want you give a first cut at the features you might be using. Make this about 1 page long. Be concise and to the point. I do not need a tutorial.

(2) (50 points) Consider a neuron with *n* inputs and one output. Let *t* be the target value of the output (i.e., the desired output or teacher) and let *y* be the actual output of the neuron. Let the output and input are related by the nonlinear activation function *f*. We can use several activation functions. The purpose of this exercise is to make you look at several activation functions and get familiar with them. For each of the following cases, (a) sketch the shape of the activation function and mark on the axes any significant information, (b) calculate the derivatives, if they exist, of each. Show your calculations.

To reduce the tedium, you may use graphing tools such as those available in Matlab, but copy the significant parts into your handout.

(a) (points 1+ 1) linear activation function, f(x) = x
(b) (points 1+ 1) the unit step function with the jump at the origin.
(c) (points 1+ 1) the step function that has a jump from –1 to +1 at the origin
(d) (points 5+ 5) the sigmoidal (or logistic) function f = 1/(1 + exp (-ax)). Plot several by changing the value of the parameter *a*
(e) (points 4+ 5) the symmetric sigmoid. If g(x) is the standard sigmoid as defined above, then a symmetric sigmoid is f(x) = 2 g(x) –1
(f) (points 5+ 5) a hyperbolic tangent function f(x) = tanh (ax)
(g) (points 5+ 5) a radial basis function, f(x) = exp (-ax**2)

(3) (40 points) Consider the Delta rule formula derived in the book on pages 91 and 92. This was derived for the minimum least squared error criterion, namely

$$E(w) = \frac{1}{2}\sum_{d \in D}(t_d - o_d)^2$$

Change this to the entropy criterion shown below and re-derive the formula and comment on the results.

$$\text{Entropy} = E = -\sum_{d \in D} t_d \log_2 o_d$$