CSCI567 Machine Learning (Fall 2008) Assignment #4

Instructor: Dr. Sofus A. Macskassy
TA: Cheol Han

Due time: 5:00pm, Nov 11, 2008

Student Name: ______________________

Student ID: ______________________

1. (**SVM**, 15 points)
The quadratic kernel $K(x_i, x_j) = (x_i \cdot x_j + 1)^2$ is equivalent to mapping each x into a higher dimensional space where
$$\Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$
for the case where $x = (x_1, x_2)$.
Now consider the cubic kernel $K(x_i, x_j) = (x_i \cdot x_j + 1)^3$.
What is the corresponding Φ function (again, for the case where $x = (x_1, x_2)$).

2. (**Bayesian Decision Theory**, 15 points) [Based on Question 3.4 in text book]
   a. Somebody tosses a fair coin and if the result is heads, you get nothing, otherwise you get $5. How much would you pay to play this game?
   b. What would a normal person (such as yourself) pay if the win is $5000 instead of $5?
   c. What if the coin comes up head 90% of the time and the win is $100?

3. (**Neural Networks,** 10 points)
Derive the update equations when the hidden units use tanh instead of the sigmoid. Use the fact that tanh' = (1-tanh^2)

4. (**VC Dimension of geometric concept classes**, 30 points)
Consider the space of instances *X* corresponding to all points in the (*x*,*y*) plane. Give the VC dimension of the following three hypothesis spaces:
   a. H_r = the set of all rectangles in the (*x*,*y*) plane. That is:
   $$H_r = \{((a < x < b) \wedge (c < y < d)) \mid a, b, c, d \in \Re\}$$
   b. H_c = circles in the (*x*,*y*) plane. Points inside the circle are classified either as positive or negative examples.
   c. H_t = triangles in the (x,y) plane. Points inside the triangle are classified as positive examples.

5. (**Error Bounds**, 30 points)
Consider the class *C* of concepts of the form $((a < x < b) \wedge (c < y < d))$, where *a, b, c* and *d* are integers in the interval [0,99]. Note that each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the (*x*,*y*) plane. Hint: Given a region in the plane bounded by (0, 0) and (*n-1, n-1),* the number of distinct rectangles with integer-valued boundaries within this region is $\left(\frac{n(n-1)}{2}\right)^2$.
   a. Give an upper bound on the number of randomly drawn training examples sufficient to assure that for any target concept *c* in *C*, any consistent learner using *H*=*C* will, with probability 95%, output a hypothesis with error ≤ 0.10.
   b. Now suppose the rectangle boundaries *a, b, c,* and *d* take on *real* values instead of integer values. Update your answer to the first part of this question.