CS534 — Homework Assignment 5 — Due Monday May 2

1. Cubic Kernels. In class, we showed that the quadratic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^2$ was equivalent to mapping each $\mathbf{x}$ into a higher dimensional space where

$$\Phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$

   for the case where $\mathbf{x} = (x_1, x_2)$. Now consider the cubic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^3$. What is the corresponding Φ function (again, for the special case where $\mathbf{x} = (x_1, x_2)$).

2. Geometry of Lines and Points. Consider the line $2x_1 + 3x_2 - 1 = 0$. What is the distance from this line to the origin (at the point where the line is closest to the origin)? Now consider a point $(x_1, x_2) = (5, 1)$. What is the distance from this point to the line?

   In general, consider the line $w_1x_1 + w_2x_2 + b = 0$. What is the general formula for the distance from this line to the origin? What is the general formula for the distance from this line to some arbitrary point (u, v)? Hint: See the textbook section 5.2.1 at page 216).

3. VC Dimension of geometric concept classes.

   Consider the space of instances X corresponding to all points in the (x, y) plane. Give the VC dimension of the following hypothesis spaces:

   (a) [5] H_r = the set of all rectangles in the (x, y) plane. That is, $H = \{((a < x < b) \wedge (c < y < d)) \mid a, b, c, d \in \Re\}$.

   (b) [4] H_c = circles in the (x, y) plane. Points inside the circle are classified as positive examples.

4. Consider the class C of concepts of the form $(a \leq x \leq b) \wedge (c \leq y \leq d)$, 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 the points $(0, 0)$ 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) [3] 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 at most 0.15.

   (b) [3] 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.