Homework 4
(due Friday, November 4th, 2005)

25th October 2005

1. A pattern $x^{(n)} \in R^D$, $n \in \{1, \ldots, N\}$ is defined as $\{x_1^{(n)}, \ldots, x_D^{(n)}\}$.

   (a) For a set of two dimensional patterns, $x^{(n)} \in R^2$, show that the polynomial kernel $k(x, y) = (x \cdot y + 1)^3$ is non-negative definite.

   (b) Also, show that the polynomial kernel $k(x, y) = (x \cdot y + 1)^d$ is non-negative definite for D dimensional patterns. You'll need to expand $(\sum_{i=1}^{D} x_i^{(m)} x_i^{(n)} + 1)^d = \sum_{d_0=0}^{D} \cdots \sum_{d_D=0}^{D} \frac{d}{\prod_{i=0}^{D} d_i!} \prod_{i=0}^{D} (p_i)^{d_i}$, where $p_0 = 1$, $p_i = x_i^{(m)} x_i^{(n)}$, $i > 0$ and $\sum_{i=0}^{D} d_i = d$. This problem is much easier than you think.

2. Work out a table of equivalents between the finite dimensional vector space and the infinite dimensional Hilbert space. Specifically, write down the equivalents for the index set, inner product, eigenvectors, eigenvalues, orthonormality condition, positive definiteness criterion, function expansion where $f = \sum_{i=1}^{N} (f^T e_i) e_i$ in the finite dimensional case, eigenvalue relation where $\Sigma e_i = \lambda_i e_i$ in the finite dimensional case, kernel function expansion where $f(x) = \sum_{i=1}^{l} \alpha_l K(x, x_l)$ in the infinite dimensional case and l the number of patterns in the training set, the RKHS property which is $< \sigma_i, \sigma_j >= \sigma_{ij}$ in the finite dimensional case and finally the inner product when the functions are expanded using a kernel as in $< f, g >$ where $f(x) = \sum_{i=1}^{l} \alpha_l K(x, x_l)$ and $g(x) = \sum_{i=1}^{l} \beta_l K(x, x_l)$ in the infinite dimensional case.

3. **Relational Clustering**: The K-means clustering algorithm can be seen as minimizing the following objective function

$$\min_{(M,\mathbf{y})} E_{\text{cluster}}(M, \mathbf{y}) = \min_{(M,\mathbf{y})} \sum_{i=1}^{N} \sum_{a=1}^{K} M_{ia} ||\mathbf{x}_i - \mathbf{y}_a||^2 \tag{1}$$

subject to the constraints $\sum_{a=1}^{K} M_{ia} = 1$ and $M_{ia} \in \{0, 1\}$ where $\{\mathbf{x}_i\}$ is the data, $\{\mathbf{y}_a\}$ the set of cluster centers and $\{M_{ia}\}$ the set of memberships of data points in clusters. We now move from point clustering to relational clustering. In the previous K-means clustering case, we assigned the membership of a data point to the current nearest cluster center. Instead of doing this, we now wish to assign the membership of data point "i" by also examining the membership of nearby data point "j" and how close "j" is to cluster "a". To do this, we first parse the data and generate a nearest neighbor graph $\{G_{ij}\}$. If $G_{ij} = 1$, it implies that "i" and "j" are neighbors. For the sake of simplicity, assume that this graph is symmetric. That is, if "i" is a neighbor of "j", then "j" is a neighbor of "i".

(a) Given this graph G, design a new *pairwise* relational clustering objective function somewhat similar to (1). The new objective function should have the following properties. i) It should only have quadratic terms linking $\{\mathbf{x}_i\}$ to $\{\mathbf{y}_a\}$ [for example $\| \mathbf{x}_i - \mathbf{y}_a \|^2$], ii) it should cover the pairs of data $\{\mathbf{x}_i\}$ which are neighbors in G, iii) it should use both the memberships of data point "i" in "a" and of data point "j" in "a" for neighbors "i" and "j" in G.

(b) Derive the solution for M_{ia} which should depend on the current value of $\mathbf{y}_a$ and the memberships M_{ja} of "i"s neighbors "j" in G.

(c) Derive the solution for $\mathbf{y}_a$ which should depend on the current value of $\{M_{ia}\}$.