CS 688 – Spring 2016

Homework 4 – Due April 25

Professor: Carlotta Domeniconi

Problem 1 Prove that, if $K_1(x, y)$ and $K_2(x, y)$ are two kernel functions, then $K(x, y) = K_1(x, y)K_2(x, y)$ is also a kernel function.

Problem 2 Consider the simple linear SVM classifier $(w_1 x + w_0)$, and the non-linear SVM classifier $(\mathbf{w}^t \phi + w_0)$, where $\phi : \Re \to \Re^2$ is defined as: $\phi(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$

(a) Provide three input points x_1, x_2, and x_3 and their associated labels (-1 or $+1$) such that they cannot be separated with the simple linear classifier, but are separable by the non-linear classifier with $\phi = (x, x^2)^t$.

(b) Mark your three points x_1, x_2, and x_3 as points in the *feature space* with their associated labels. Draw the decision boundary of the non-linear SVM classifier that separates the points *in the feature space* obtained with $\phi = (x, x^2)^t$.

Problem 3 (**Clustering**) Consider the 3-means algorithm on a set S consisting of the following six two-dimensional points: $a = (0, 0)$, $b = (8, 0)$, $c = (16, 0)$, $d = (0, 6)$, $e = (8, 6)$, $f = (16, 6)$. The algorithm uses the Euclidean distance to assign each point to the nearest centroid; ties are broken in favor of the centroid to the left/down. A starting configuration is a subset of three starting points from S that form the initial centroids. A 3-partition is a partition of S into 3 subsets; thus, {a,b,e}, {c,d}, {f} is a 3-partition. Clearly, any 3-partition induces a set of three centroids in a natural way. A 3-partition is stable if repetition of the 3-means iteration with the induced centroids leaves it unchanged.

1. How many starting configurations are there?
2. What are the stable 3-partitions?
3. What is the number of starting configurations leading to each of the stable 3-partitions.
4. What is the maximum number of iterations from any starting configuration to its stable 3-partition?

Problem 4 (**Clustering**) Implement the K-means algorithm, and run it on the data given above to verify your answers to Problem 3. You need to provide the output of K-means (initial configuration and corresponding stable 3-partition) that shows all your testing scenarios.

Instructions Complete the homework by April 25. Turn in the source code electronically as an attachment by email to carlotta@cs.gmu.edu. Make sure your code compiles and runs properly. **Turn in also a hardcopy of your homework (including a copy of the source code and testing scenarios) in class on April 25**.

Important I am well aware that the source code of the k-means algorithm is available online. **Submission of a code that is not your own is a violation of the honor code, and has serious consequences**.