

Name: .....

Student ID#: .....

**Statistical Pattern Recognition (CE-725)**  
**Department of Computer Engineering**  
**Mini Exam #1 (SVM & Kernel) - Spring 2012**

**1. (15 points)** Prove that  $K(x, x') = \exp(-\|x - x'\|^2 / c)$  is a valid kernel. You can only use the following given identities:

If  $k_1(x, x')$  and  $k_2(x, x')$  are valid kernels, then the following  $k(x, x')$ s are also valid kernels:

- |   |                                       |
|---|---------------------------------------|
| 1. $k(x, x') = ck_1(x, x')$                 | 4. $k(x, x') = \exp(k_1(x, x'))$      |
| 2. $k(x, x') = k_1(x, x') + k_2(x, x')$     | 5. $k(x, x') = f(x) k_1(x, x') f(x')$ |
| 3. $k(x, x') = k_1(x, x') \cdot k_2(x, x')$ |                                       |

**2. (15 points)** Consider the training points as below:

$$C_1 = \{ (1,3)^T, (1,4)^T, (2,2)^T \}; \quad C_2 = \{ (4,0)^T, (5,0)^T, (6,1)^T \}$$

Plot data points and draw the large margin classifier that SVM finds visually. Then without explicitly solving the optimization problem, determine support vectors and  $\lambda$ s for each training points.

**3. (15 points)** Suppose that you are given n-dimensional real-valued feature vectors. You are thinking of using as kernel a function that counts in how many places both  $x$  and  $x'$  are greater than some constant threshold  $\theta$ . In other words, let:

$$\delta_\theta(a, b) = \begin{cases} 1 & \text{if } a > \theta \text{ \& } b > \theta \\ 0 & \text{otherwise} \end{cases}$$

Then the function defined by  $k_\theta(x, x') = \sum_{j=1}^n \delta_\theta(x_j, x'_j)$  and  $x_j$  denotes  $j^{th}$  component of  $x$ . Show that this function is a valid kernel.

**4. (20 points)** After mapping into higher dimensional feature space, through a radial basis (Gaussian) kernel function, is 1-NN using unweighted Euclidean distance able to achieve better classification performance than in the original space?

**5. (15 points)** Suppose we run SVM on samples with features  $x_1, \dots, x_n$  i.e.  $x = (x_1, \dots, x_n)^T$ , and obtain the boundary. Then, we add a random feature  $x_{n+1}$  to feature vectors. How does the boundary changes by adding this feature?

**6. (20 points)** Consider the following classification problem. We will use soft margin linear SVM for this problem. Use the following figures to show where the decision boundaries are likely to be for  $c=0.05$  (in the left figure) and  $c=10000$  (in the right figure), respectively. Don't solve the problem explicitly; just give a brief justification for each case.

