Homework Assignment #6  
Due: 04/16/2010 (11:55pm, through Oncourse)

**Question 1** (50 points)
Modify your K-NN predictor (function knn.m from the class web site can be used) designed for classification to allow for a variable number of input parameters. The function should allow for the following types of function calls

```plaintext
prediction = knn(D, T);
prediction = knn(D, T, 'distance', 'euclidean');
prediction = knn(D, T, 'K', 5);
prediction = knn(D, T, 'distance', 'mahattan', 'K', 3);
prediction = knn(D, T, 'K', 7, 'distance', 'lmax');
```

The following distance measures should be supported as strings, after the string ‘distance’: ‘euclidean’, ‘mahattan’ and ‘lmax’. For parameter $K$, any positive integer value should be accepted (make sure you cover the case when $K$ is greater than the number of training data points). Your program should also carry out appropriate testing of input parameters such that only strings are accepted for the distance measure specification and that only integer numbers are accepted for $K$. Also, only strings should be allowed that specify type of input (e.g. ‘$K$’ vs. ‘distance’). The default value for $K$ should be 1 and the default for the distance measure should be ‘euclidean’ (the defaults are applied if the user does not specify a particular option). $D$ and $T$ are specified as in previous assignments (training and test set, respectively).

**Extra Credit Question** (25 points)
Modify your K-NN predictor (function knn.m from the class web site can be used) designed for classification to have the following form

```plaintext
function [prediction] = knn (D, T, K, distance_function)
```

where $D$, $T$ and $K$ are specified as before, while `distance_function` is a function handle specified by an external user.

For example, the function can be called in the following way

```plaintext
D = load('some_training_data.txt');
T = load('some_test_data.txt');
K = 5;
distance_M = (@(x, y) sum(abs(x - y)));
p = knn(D, T, K, distance_M);
```

**What to Submit?** Two Matlab files, named knn.m and knn_extra.m.
Homework Policy

Homework assignments are to be submitted through Oncourse on or before the specified due date. Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rule:

- on time: your score × 1
- up to 1 day late: your score × 0.9
- up to 2 days late: your score × 0.7
- up to 3 days late: your score × 0.5
- up to 4 days late: your score × 0.3
- up to 5 days late: your score × 0.1

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points (that is, 50% of the credit you receive for your answers).

All assignments are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Good Luck!