**Intelligent Data Analysis**

**Problem Set 6** **Due 5/26/2016**

1. This problem is about local linear regression and equivalent kernel discussed in Chapter 11, pp 30-33.

   (a) Show details of the derivation of the optimal parameters

   $$\begin{bmatrix} w_0 \\ \mathbf{w} \end{bmatrix} = (\mathbf{X}^T\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{K}\mathbf{r}, \text{ where } \mathbf{X} = \begin{bmatrix} 1 & (\mathbf{x}^1)^T \\ \vdots & \vdots \\ 1 & (\mathbf{x}^N)^T \end{bmatrix} \text{ and}$$

   $$\mathbf{K} = \begin{bmatrix} K\left(\dfrac{\mathbf{x}-\mathbf{x}^1}{h}\right) & & \\ & \ddots & \\ & & K\left(\dfrac{\mathbf{x}-\mathbf{x}^N}{h}\right) \end{bmatrix}.$$

   (b) Define the equivalent kernel $I^t(\mathbf{x})$ as the *t*th component of the *N*-dimensional vector $\mathbf{I}(\mathbf{x}) = \mathbf{K}\mathbf{X}(\mathbf{X}^T\mathbf{K}\mathbf{X})^{-1}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}$. Use $\mathbf{X}^T\mathbf{I}(\mathbf{x}) = \tilde{\mathbf{x}}$ to show that $\sum_{t=1}^{N} I^t(\mathbf{x}) = 1$ and $\sum_{t=1}^{N} I^t(\mathbf{x})(\mathbf{x}^t - \mathbf{x}) = \mathbf{0}$.

   (c) If the local linear regression $g(\mathbf{x}) = w_0 + \mathbf{w}^T\mathbf{x}$ is replaced by the "reduced model" $g(\mathbf{x}) = w_0$, follow the procedure in pp 30-31 to show that the optimal "reduced model" is equivalent to kernel weighted smoother (pp 28):

   $$g(\mathbf{x}) = \frac{\sum_{t=1}^{N} K\left(\dfrac{\mathbf{x}-\mathbf{x}^t}{h}\right) r^t}{\sum_{t=1}^{N} K\left(\dfrac{\mathbf{x}-\mathbf{x}^t}{h}\right)}.$$

2. Program Assignment In this assignment, you are asked to do classification using Fisher's iris data set, which contains 50 specimens from each of three different species of iris—setosa, versicolor, and virginica—on the following dimensions: sepal length, sepal width, petal length, petal width. Use the first 100 cases as the training and the rest 50 cases as the testing data.

   You are asked to implement the following prototype method and nonparametric classifiers with the training data. Each classifier uses all the four input variables and returns the predicted class label. You may use Euclidean distance as the metric.

(a) Prototype method using $K$-means clustering, each class with 5 prototypes (see Chapter 10, pp 4).

(b) Prototype method using LVQ, each class with 5 prototypes (see Chapter 10, pp 12).

(c) Kernel density classifier (see Chapter 11, pp 14) with multivariate Gaussian kernels $K\left(\dfrac{\mathbf{x}-\mathbf{x}^t}{h}\right) = K(\mathbf{u}) = \left(\dfrac{1}{\sqrt{2\pi}}\right)^d \exp\left[-\dfrac{\|\mathbf{u}\|^2}{2}\right]$, where $d$=4. Try as least three different values of $h$.

(d) Distance weighted $k$-nearest neighbors with $k$ = 1, 3, 5 (see Chapter 11, pp 18). You may use the same kernel as (c). Try different values of $h$.

Apply the 50 testing cases to each classifier, report their classification matrix and compute the accuracy. Report any interesting findings you got.