Homework 2
Stat 215A, Fall 2014

Due: Friday October 7, 2013, 4:00 PM
Please submit to bCourses and hand in a hard copy along with your lab 2

1 Kernel density estimation

Explain in your own words the bias variance trade-off in histograms and in kernel density estimates. Write down the bias term of the mean squared error of the kernel density estimator,

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x)$$

at the point x. Does the bias change with n? with h? Does this make sense in the context of your previous answer?

2 Multidimensional scaling

For the following questions, use the MDS reading in Mardia et al. [1980] in Lab 2 under bSpace. For a distance matrix $\mathbf{D}$, let,

$$\mathbf{A} = (a_{rs}), \text{ where } a_{rs} = -\frac{1}{2} d_{rs}^2$$

and set

$$\mathbf{B} = \mathbf{HAH}$$

where $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$.

1. Show if $\mathbf{D}$ is the matrix of Euclidean interpoint distances for a configuration $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)'$ then

   $$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})'(\mathbf{z}_s - \bar{\mathbf{z}}) \quad r, s = 1, \ldots, n$$

   Show that this implies $\mathbf{B}$ is positive semidefinite.

2. Let $\mathbf{B}$ be positive semidefinite of rank p with positive eigenvalues $\lambda_1 > \ldots > \lambda_p$ and corresponding eigenvectors $\mathbf{X} = \left(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(p)}\right)$ normalized such that

   $$\mathbf{x}_{(i)}'\mathbf{x}_{(i)} = \lambda_i \quad i = 1, \ldots, p$$

   (a) Show that the points with coordinates $\mathbf{x}_r = (x_{r1}, \ldots, x_{rp})'$ have interpoint distances given by $\mathbf{D}$.

   (b) Further show that the configuration has center of gravity $\bar{\mathbf{x}} = \mathbf{0}$ and $\mathbf{B}$ represents the inner product matrix for this configuration.

References

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis.* Academic Press, London, 1980.