University of Colorado
Department of Civil, Environmental and Architectural Engineering
Advanced Data Analysis Techniques
CVEN 6833

Homework Set 1

Date :09/12/2005
Due :10/01/2005

1. Develop an R code to perform local polynomial method. Use the synthetic data from the hand out to test this. Specifically evaluate the
   i. L matrix
   ii. GCV
   iii. Estimate and compare with those from a linear regression approach.

2. Piezzometric heads during 1998 at 62 wells from the Bangkok aquifer system can be obtained from http://civil.colorado.edu/~balajir/CVEN-6833/bangkok-1998-welldata The first column is the head values, the X and Y co-ordinates of the well locations are in the second and third columns, respectively. As a water manager on this aquifer system you are interested in obtaining the underlying piezzometric surface from these sparse spatial observations – as this will be used in estimating the travel time of contaminants through the system.
   In order to estimate the surface you would like to try a parametric (i.e. linear regression) and nonparametric (i.e. LOCFIT) approach.
   (Make one run with your LOCFIT that you developed in 1. above and compare with the commands from the LOCFIT library)
   i. Fit the models. For LOCFIT obtain the best alpha and degree of polynomial using GCV or CV. Use the best values to make estimates at the observation points.
   ii. Fit a linear regression model.
   ii. Perform diagnostics on both the models and ANOVA.
   iii. Estimate the piezzometric heads on a 5 x 5 grid with the ranges of X co-ordinate to be between 42.5 km and 117.5 km; and that of Y to be between -152.5km and -67.5 km. Make surface plot of the estimates. You can compare this to the surface plot of the raw data.
   iv. Compute the cross-validated estimates of the raw data and surface plot them as in ii.
   v. Estimate and surface plot the standard error (i.e. prediction error).
   vi. Also obtain the confidence intervals from bootstrap.
   vii. You may also withhold part of the data and do a prediction on the data withheld and evaluate the performance.
   Comment on the results from these two methods.