Regression Density Estimation and Stochastic Approximation

In Nott, Tan, Villani and Kohn (2012) (Journal of Computational and Graphical Statistics, 21 (3), 797-820) we consider the problem of flexibly estimating a response distribution in regression as a function of covariates. An important approach to regression density estimation uses finite mixture models and our article considers flexible mixtures of heteroscedastic regression (MHR) models where the response distribution is a normal mixture, with the component means, variances and mixture weights all varying as a function of covariates. Our article develops fast variational approximation methods for inference. Our motivation is that alternative computationally intensive Monte Carlo methods for fitting mixture models are difficult to apply when it is desired to fit models repeatedly in exploratory analysis and model choice.

Figure 1 below illustrates the results of applying our method for a four component mixture model. The response to be predicted is the output of a deterministic hydrological streamflow model, to be predicted as a function of two input parameters. The mixture approach we use follows a "divide and conquer" approach to modelling, effectively softly splitting the data into pieces, with regression surfaces fitted to each piece. The rows of the figure below show the estimated means and standard deviations for each piece, computed using our new highly efficient variational approximation method which is able to handle very large data sets.

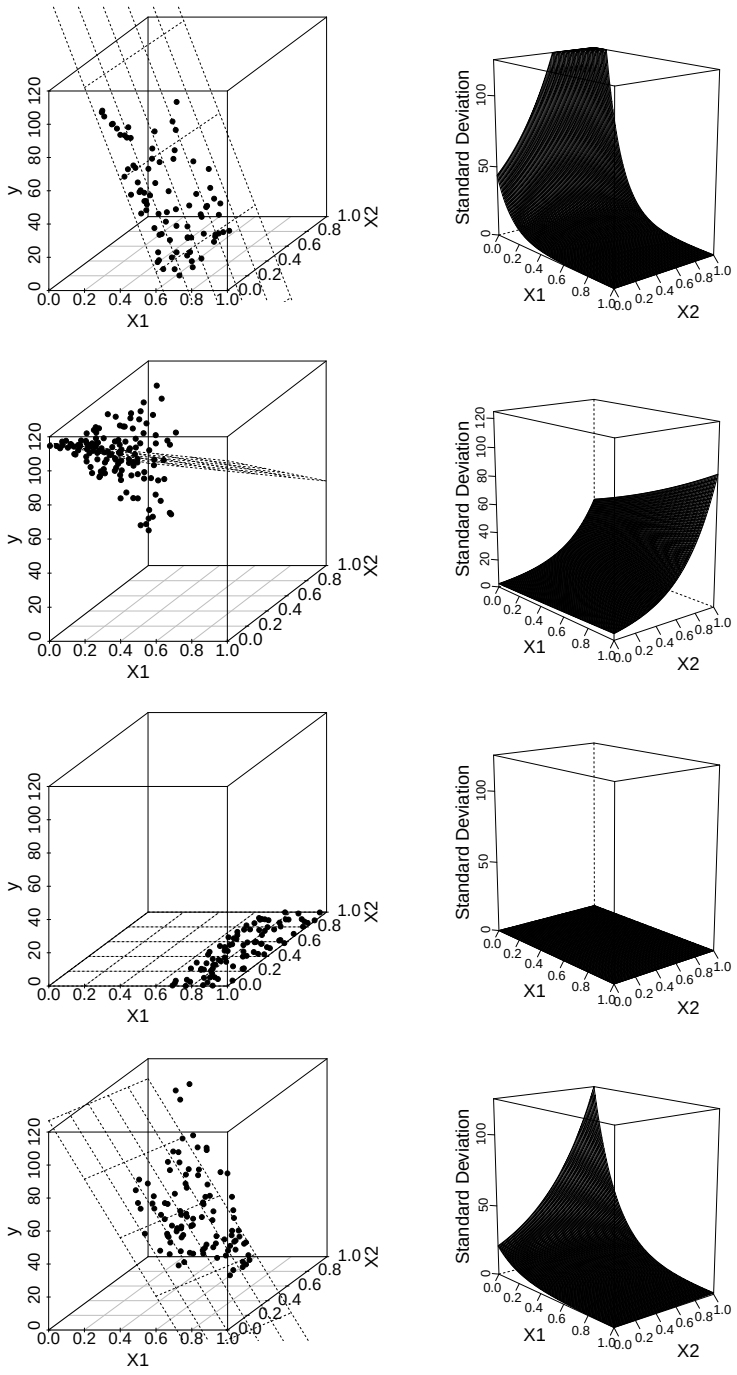

Figure 1: Fitted component means (first column) and standard deviations (second column) for four component mixture model for rainfall-runoff example.