Localized Sliced Inverse Regression

Code

LSIR.m is the code. The Tai-Chi example was generated by running ex_digits_lsir_lin.m.

Inputs and Outputs

The inputs of this function should consist of a covariate matrix X with columns corresponding to observations and rows dimensions (input variables), a response vector Y with length equal the number of columns of X, the number of effective dimension reduction (EDR) dimensions d, a regularization parameter s, and (optionally) a structure type variable that specifies performing regression or classification, the number of slices and the number of nearest neighbors. The function will return a structure variable that contains the estimated EDR directions and some other quantities. Try "help LSIR" in the command line for more details.

Examples

Tai-Chi

In this section we illustrate how dimension reduction is performed by considering a classification problem for the Tai-Chi data. The covariate matrix X has six rows (hence six dimensions). An illustration of the first two dimensions is shown in Figure 1(a). The third to sixth dimensions are independent random errors. The number of columns (the sample size) is taken to be 1000 (500 for each of the red and blue points). The response Y is simply a vector of -1 and 1 with -1 corresponding to the red points and 1 the blue points.

Now obviously the true EDRs should be (e1, e2) where e1=(1,0,0,0,0,0)' and e2=(0,1,0,0,0,0)'. One can type:

[LS] = LSIR(X, Y, 2, 0, opts);

in the command line with opts.pType = 'c' and opts.numNN = 10. The third argument tells the function to choose 2 EDRs, and the fourth argument sets the regularization parameter to be 0. opts is a structure: opts.pType='c' means performing classification and opts.numNN=10 means the number of nearest neighbors is taken to be 10.

The output LS is a structure, and usually LS.edrs which represents the estimated EDRs is of concern. By typing "plot(LS.edrs(:,1)'*X(:,Y==1) , LS.edrs(:,2)'*X(:,Y==1))" followed by "hold on; plot(LS.edrs(:,1)'*X(:,Y==-1) , LS.edrs(:,2)'*X(:,Y==-1))" one projects the training data X onto the estimated 2 EDRs and visualizes it. The resulting figure is shown in Figure 1(b). We also form an independent test dataset formatted like X and repeat the project and visualize procedure, and the resulting figure is shown in Figure 1(c).

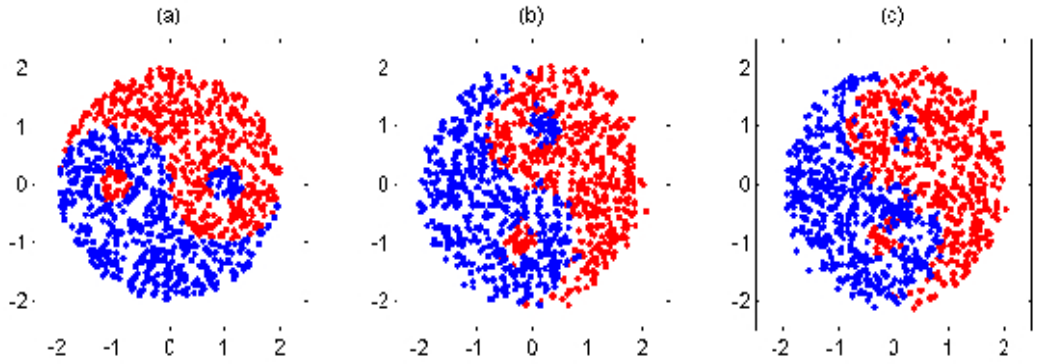

Figure1: The first two dimensions of the Tai-Chi data (a), projection (on the estimated first two EDRs) of the training data (b) and projection (on the estimated first two EDRs) of an independent test data (c).

For any questions and comments please email km68 at stat dot duke dot edu. Thanks.