

Description of the HDR library

Yilu Zhang

1 Introduction

Hierarchical discriminant regression (HDR) is a new hierarchical statistical modeling method. It has been proved to be a powerful tool for applications involving function approximation and classification.

HDR casts classification problems (class labels as output) and regression problems (numeric values as output) into a unified regression problem. This view enables classification problems to use numeric information in the output space which is available for regression problems but is traditionally not readily available for classification problems. HDR does clustering in both input and output space. Discriminant analysis is enforced because the clustering in input space is guided by the clustering in output space. Both clusterings are done in a coarse-to-fine manner so as to reduce the computational complexity.

HDR has been applied to the problems such as face recognition, OCR, vision-based navigation, speech recognition, and stock prediction problems. Comparing to existing techniques such as decision tree, artificial neural networks, HDR has superior performance in handling high dimensionality and conducting one-instance learning.

2 HDR library

There are two modes to implement HDR, the batch mode and the incremental mode. Enclosed in this package is the batch mode library. There are two major subroutines offered in the batch mode HDR library.

```
void BuildTree(int nType,char *trfile,int km,int xdim,int verbose,float alpha,float bound,char *treename);

void TestTree(int nType,char *tefile,int km,int xdim,int verbose,float alpha,char *treename);
```

where,

nType: the type of training file or testing file.

- nType=0: The training file (or testing file) includes the data, with each line corresponding to one data followed by its label.
- nType=1: The training file (or testing file) includes the file names of the images followed by their labels. Each file name corresponds to one image of .pgm format.

- nType=2: The training file (or testing file) includes the file names of the images followed by their labels. Each file name corresponds to one image of raw format.
- Note: For both nType=0 and nType=2, the user needs to specify the data dimension, xdim.

trfile: the training file

tefile: the testing file

km: the number of clusters in a node of the HDR tree

xdim: the dimension of input data

verbose: if verbose==1, then report more than just predicting the label for the testing examples

alpha: the switch confidence (default value 0.1), which typically does not need to change

bound: accuracy parameter (range: 1.0-2.0)

treename: the name of the file to save the tree

The HDR library is generated with the UNIX `ar` utility, which combines the object files of the library into a single archive file, `hdr.ar`. One needs to add the switch `-a` when using `CC` to link the library. A sample makefile is given to shown how to do this.

3 Also included

A sample application of the HDR library is given under subdirectory `sample_codes` together with a makefile. A few face images from FERET face dataset (raw format 88*64) are given under subdirectory `data` for the sample application codes.