Approximate Information Discriminant Analysis (AIDA)

Koel Das

February 11, 2008

1 Introduction

This tutorial is an accompanying document to the computer code for *approximate information discriminant analysis*. The details of the method can be found in [1] and the computer code is written in MATLAB^{TM}. The code consists of several basic functions:

(1) `aida.m`

(2) `sqrtmat.m`

(3) `invsqrtmat.m`

(4) `logmat.m`

`aida.m` returns the optimal feature extraction matrix, T^*. The input arguments of `aida.m` allow for choice of size of feature space. Type help `aida` in MATLAB^{TM}command prompt to learn more about this function.

`sqrtmat.m` is a function that extracts squareroot of any covariance matrix `A`, as defined in [1]. This function is an auxiliary function in `aida.m`.

`invsqrtmat.m` is a function that extracts the squareroot of inverse of any covariance matrix, `A` as defined in [1]. This function is an auxiliary function in `aida.m`.

`logmat.m` is a function that extracts the logarithm of any covariance matrix `A`, as defined in [1]. This function is an auxiliary function in `aida.m`.

2 Example

Application of these functions is illustrated on data set Wine from the UCI machine learning repository. This data set consists of 178 samples (each sample is 13-D). The number of classes is 3.

Fig. 1 shows the resulting 2-D feature plot taking all the 178 samples. For plotting the Wine feature space, the function `plot_wine.m` is posted. Note that the function is specifically tailored to the Wine data. Also note that it is assumed that a file `wine.mat` resides in the same directory as `plot_wine.m`. This file should contain a variable `Data` in the form of 178×14 matrix. The last column of this matrix contains class label indicators (integers), e.g. $\{0, 1, 2\}$. Type `help plot_wine` or take a look at the source code for further information on this function.

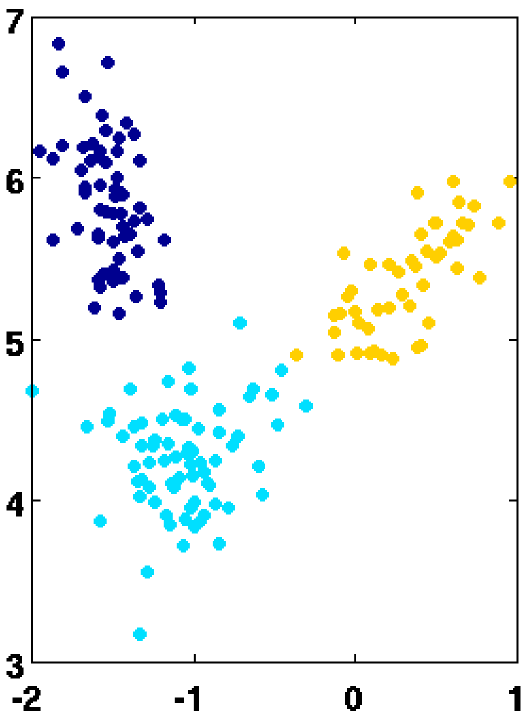

Figure 1: 2-D features corresponding to the samples of the Wine data. Colors indicate class memberships.

For example, running `[T et] = plot_wine(2,1);` from MATLAB™command prompt returns the following:

- The optimal feature extraction matrix `T`.
- The elapsed time `et`.

The input arguments represent:

- The size of the feature space `m = 2` (passed to (1)).
- Plot indicator `PlotFlag = 1` to enable the plotting of features.

Table 1: The performances (%) estimated through leave-one-out CV using linear and quadratic classifiers for dataset Wine

Dataset	m	1	3	5	13 (full)
Wine	AIDA(L)	91.57	$\boxed{99.44}$	98.88	98.88
	ACC(L)	89.89	98.88	98.31	98.88
	MLT(L)	82.58	97.75	–	98.88
	Tubbs(L)	67.41	74.72	91.57	98.88
	LDA(L)	91.01	98.88	–	98.88
	LPP(L)	92.13	98.31	98.31	98.88
	AIDA(Q)	92.13	99.44	$\boxed{100}$	99.44
	ACC(Q)	91.01	99.44	99.44	99.44
	MLT(Q)	85.95	99.44	–	99.44
	Tubbs(Q)	70.22	76.4	97.19	99.44
	LDA(Q)	91.57	99.44	–	99.44
	LPP(Q)	92.13	98.31	99.44	99.44

The performance of AIDA criterion was tested experimentally with the *Wine* dataset and compared to the performances of LDA, ACC [3], Tubbs [4], MLT [5] and LPP [6]. The performance was evaluated using linear and quadratic Bayesian classifiers, with prior probabilities estimated empirically from the data and the Gaussian approximation of the class-conditional probability density functions. A leave-one-out cross validation (CV) was used for performance evaluation. The admissible subspace dimensions are $1 \leq m \leq c-1$ for LDA, $1 \leq m \leq \frac{1}{2}c(c-1)$ for MLT [5], and $1 \leq m \leq n-1$ for the AIDA, Tubbs, LPP and ACC methods. Table 1 shows results, represented by the estimated correct classification rates, for various dimensions of the feature space. The boxed values represent the best results for each classifier.

References

[1] K. Das and Z. Nenadic, Approximate Information Discriminant Analysis: a Computationally Simple Heteroscedastic Feature Extraction Technique, *Pattern Recogn.*, vol. 41 (5), pp. 1565-1574, 2008.

[2] Z. Nenadic, Information discriminant analysis: Feature extraction with an information-theoretic objective, *IEEE T. Pattern Anal.*, vol. 29 (8), pp. 1394-1407, 2007.

[3] M. Loog and R.P.W. Duin, Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion, *IEEE T. Pattern Anal.*, vol. 26, pp. 732-739, 2004.

[4] J.D. Tubbs, W.A. Coberly, D.M. Young, Linear dimension reduction and Bayes classification with unknown population parameters, *Pattern Recogn.*, vol. 15(3), pp. 167-172, 1982.

[5] H. Brunzell and J. Eriksson, Feature reduction for classification of 19 multidimensional data, *Pattern Recogn.*, vol. 33(10), pp. 1741-1748, 2000.

[6] X. He, S. Yan, Y. Hu, P. Niyogi, Face recognition using Laplacian faces, *IEEE T. Pattern Anal.*, vol. 27(3), pp. 328-340, 2005.