

JCU ePrints

This file is part of the following reference:

Mallet, Yvette Lelia (1997) *Wavelet based feature extraction methods for the discrimination and regression of spectral data*. PhD thesis, James Cook University.

Access to this file is available from:

<http://eprints.jcu.edu.au/17437>



WAVELET BASED
FEATURE EXTRACTION METHODS
FOR THE
DISCRIMINATION AND REGRESSION
OF
SPECTRAL DATA

Thesis submitted by
Yvette Lelia MALLET Bsc(Hons) *Qld*
in October 1997

for the degree of Doctor of Philosophy
in the School of Computer Science, Mathematics and Physics
James Cook University of North Queensland

IN MEMORY OF TES EVERINGHAM

STATEMENT OF ACCESS

I, the undersigned, the author of this thesis, understand that James Cook University of North Queensland will make it available for use within the University Library and, by microfilm or other means, allow access to users in other approved libraries. All users consulting this thesis will have to sign the following statement:

In consulting this thesis I agree not to copy or closely paraphrase it in whole or in part without the written consent of the author; and to make proper public written acknowledgement for any assistance which I have obtained from it.

Beyond this, I do not wish to place any restriction on access to this thesis.

.....
3/3/98
.....
(Date)

STATEMENT ON SOURCES
DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

.....
3/3/06
(Date)

ACKNOWLEDGEMENTS

Foremostly, I thank my supervisors A/Prof Danny Coomans and A/Prof Olivier de Vel, for their professionalism and encouragement in the development of this thesis. I am also grateful for the many hours that they have spent revising this and other manuscripts throughout the course of my research work.

I extend my appreciation to Professor Massart for allowing me to visit his laboratory at the Free University of Brussels and providing me with the opportunity to become more familiar with NIR data. Also from the Free University, I thank Delphine Jouan-Rimbaud, Paula Fernandez, Eric Bouveresse, Wu Wen, Beate Walczak and Wim Penninckx, for their assistance.

Sincere thanks is expressed to Dr Jaroslav Kautsky at Flinders University in Adelaide, who gave up his valuable time to introduce me to wavelets. I would also like to thank Dr Bill Moran, Radka Turcajová and Pavel Turcaj for their assistance and cooperation whilst I was visiting Flinders University.

I extend my appreciation, to Professor Trevor Hastie for supplying his S-plus code and providing valuable input. Many people have provided data which I have used as part of my thesis, these people are acknowledged in Sections 7.2 and 8.2 of this thesis.

Thanks to my colleagues in the School of Computer Science, Mathematics and Physics at James Cook University for their encouragement during the final stages of this thesis. In particular to Dr Wayne Read for revising parts of this thesis. The encouragement provided by A/Prof Bob Staudte was also much appreciated.

Whilst pursuing the research summarized in this thesis, I was primarily supported by the Australian Government in the form of an Australian Postgraduate Research Award.

I wish to thank my fiancé, Andrew Everingham for his continual support and encouragement. I would also like to thank my team-mates for their patience and understanding. Finally, I thank my mother for always thinking of the little things that mean so much.

ABSTRACT

This thesis is concerned with the application of statistical methods to spectral data. A major concern which arises from spectral data is that the number of variables or dimensionality usually exceeds the number of available spectra. This leads to a degradation in performance of traditional statistical methods. There are basically two strategies which can be implemented for overcoming such situations. It is common practice to first reduce the dimensionality of the data by some feature extraction preprocessing method, and then use an appropriate low dimensional statistical procedure. An alternative procedure is to use a high dimensional statistical procedure which is capable of handling a large number of variables. This thesis considers both approaches, and investigates the applicability of wavelets as features for statistical analyses, as well as other feature extraction procedures. The particular statistical analyses investigated are discriminant and regression analysis.

It is shown that, the wavelet based methods, particularly wavelets which have been designed to suit a particular task, perform quite adequately when compared to traditional approaches.

Contents

1	Thesis Summary	1
1.1	Overview	1
1.2	Thesis Structure and Contribution	9
2	Discriminant Analysis	12
2.1	Introduction	12
2.2	Notation	16
2.3	Fisher’s linear Discriminant Analysis (FLDA)	17
2.4	Flexible Discriminant Analysis (FDA)	19
2.5	Penalized Discriminant Analysis (PDA)	26
2.6	Bayesian Classifiers	27
2.6.1	Bayesian Linear Discriminant Analysis (BLDA)	28
2.6.2	Bayesian Quadratic Discriminant Analysis (BQDA)	29
2.7	Regularized Discriminant Analysis (RDA)	29
2.8	Assessment of Model Performance	31
2.8.1	Assessment Criteria	31
2.8.2	Choosing the Evaluation Set	33
3	Regression Analysis	37
3.1	Introduction	37
3.2	Notation	38
3.3	Multiple Linear Regression (MLR)	39
3.4	Principal Component Regression	40

3.5	Partial Least Squares Regression	40
3.6	Assessment of Model Performance	42
3.6.1	Assessment Criteria	42
3.6.2	Choosing the Evaluation Set	44
4	Feature Extraction	47
4.1	Feature Selection	48
4.1.1	Feature Selection Strategies for Discriminant Analysis	48
4.1.2	Feature Selection Strategies for Regression Analysis	50
4.1.3	Classification and Regression Trees (CART)	53
4.2	Feature Transformation	55
4.2.1	Preprocessing Methods and Transformations	55
4.2.2	Principal Component Analysis (PCA)	63
4.2.3	Fourier Transform (FT)	66
4.2.4	Discrete Wavelet Transform (DWT)	67
5	Wavelets	70
5.1	Introduction	72
5.2	Fourier Transform	74
5.3	Windowed Fourier Transform	74
5.4	Continuous Wavelet Transform	75
5.5	Discrete Wavelet Transform	76
5.6	Multiresolution Analysis	77
5.7	Fast Wavelet Transform	81
5.8	Higher Multiplicity Wavelets	82
5.9	The Discrete Wavelet Transform of Discrete Data	84
5.10	The m -band Discrete Wavelet Transform of Discrete Data	96
5.11	The m -Band Discrete Wavelet Transform of a Discrete Data Set	98
5.12	Filter Coefficient Conditions	100

5.13	Boundary Related Issues	102
5.14	The Wavelet Packet Transform of Discrete Data	103
5.14.1	The Best Basis Algorithm	105
5.14.2	The Local Discriminant Basis Algorithm	107
6	Adaptive Wavelets	109
6.1	Introduction	109
6.2	Factorization of Wavelet Matrices	110
6.3	Criteria Measures for Optimization	113
6.3.1	Discriminant Criterion Functions	113
6.3.2	Regression Criterion Functions	115
6.4	The Adaptive Wavelet Algorithm	116
6.5	Example	118
7	Classification Applications	121
7.1	Overview	121
7.2	The Data Sets	122
7.2.1	Seagrass Data	122
7.2.2	Mineral Data	123
7.2.3	Paraxylene Data	124
7.2.4	Butanol Data	125
7.3	Discriminant Analysis Based on the Original Variables	126
7.4	Discriminant Analysis Based on Wavelet Coefficients	131
7.4.1	Exploring the DWT	132
7.4.2	Banded Discriminant Analysis	133
7.4.3	Stepwise Feature Extraction from the DWT	139
7.4.4	Local Discriminant Bases	143
7.4.5	Adaptive Wavelet Algorithm	145
7.4.6	Summary of the Wavelet Feature Extraction Strategies	151

7.5	Which Classification Strategy?	153
7.5.1	Performance Based Measures	154
7.5.2	Qualitative Assessment	156
7.6	Summary	168
8	Regression Applications	170
8.1	Overview	170
8.2	The Data Sets	171
8.2.1	Sugar Data	171
8.2.2	Wheat Data	172
8.3	Common Approaches for the Regression of Spectral Data	173
8.4	Regression Analysis Using Features From the DWT	176
8.4.1	Exploring the DWT	178
8.4.2	Banded Multiple Linear Regression (BMLR)	178
8.4.3	Stepwise Feature Extraction	181
8.4.4	Adaptive Wavelet Algorithm	184
8.4.5	Summary of Wavelet Based Feature Extraction Strategies	187
8.5	Which Regression Strategy?	188
8.5.1	Performance Based Measures	189
8.5.2	Qualitative Assessment	196
8.6	Summary	211
9	Concluding Remarks	213
9.1	Original Contribution	213
9.2	Summary of Results	214
9.3	Future Work and General Remarks About the AWA	217
A		219

List of Figures

1.1	A spectrum obtained from a sample of paraxylene.	2
1.2	The electromagnetic spectrum.	2
1.3	A discriminant analysis problem.	3
1.4	Feature extraction model.	5
1.5	Some wavelet basis functions.	7
1.6	Integrated feature extraction model.	8
1.7	Thesis outline.	9
2.1	Percentage of correctly classified objects obtained by three discriminant techniques (D1,D2 and D3) for eight combinations of dimensionality and class sample sizes.	13
2.2	Summary of some discriminant analysis methods.	15
2.3	A scatterplot of the discriminant scores produced by FLDA.	19
2.4	The FDA algorithm.	20
3.1	Partial least squares algorithm.	42
4.1	A CART model.	53
4.2	Demonstration of the SNV transformation.	56
4.3	Demonstration of detrending combined with the SNV transformation.	58
4.4	Demonstration of the hull quotient.	59
4.5	Demonstration of the second derivative transformation.	60

4.6	A simplified procedure for performing the second derivative transformation.	61
4.7	Demonstration of mean centering.	63
5.1	Fourier and wavelet coefficient of a sampled sine signal, with (right) and without (left) a small disturbance.	71
5.2	Some wavelet basis functions from the Daubechies family.	76
5.3	Pictorial representation of a 2 band DWT for a signal which has been sampled 8 times.	90
5.4	Labelling of the bands in the DWT.	91
5.5	2-band DWT performed on a generated spectrum to level three.	92
5.6	Another presentation for a 2-band DWT performed on the generated spectrum to level three.	93
5.7	Two-band DWT for a spectrum to six levels.	95
5.8	A 3-band discrete wavelet transform.	97
5.9	Boxplots obtained from the correlation coefficients discussed for Table 5.1.	100
5.10	Wavelet packet transform with $m = 2$	104
5.11	Best basis algorithm.	106
5.12	Best basis.	107
6.1	The adaptive wavelet algorithm.	117
7.1	Five sample spectra from the seagrass data.	123
7.2	Five sample spectra from the mineral data.	124
7.3	Five sample spectra from the paraxylene data.	125
7.4	Five sample spectra from the butanol data.	126
7.5	Correct classification rates (CCR) and quadratic probability measures (QPM) for the seagrass (s), mineral (m), paraxylene (p) and butanol (b) data.	131
7.6	The DWT and inverse DWT performed on the seagrass data.	134

7.7	The DWT and inverse DWT performed on the mineral data. . . .	135
7.8	The DWT and inverse DWT performed on the paraxylene data. .	136
7.9	The DWT and inverse DWT performed on the butanol data. . . .	137
7.10	Coefficients selected from the DWT by SWBLDA.	142
7.11	Selected wavelet coefficients (asterisks) from the best bases. . . .	144
7.12	Discriminant measure versus iteration for the adaptive wavelet algorithm.	149
7.13	Correct classification rates (CCR) and quadratic probability mea- sures (QPM) for the wavelet based methods applied to the sea- grass (s), mineral (m), paraxylene (p) and butanol (b) data. . . .	152
7.14	Correct classification rates for each of the discriminant strategies.	155
7.15	Wavelengths selected by SBLDA, SBQDA and FDA.	158
7.16	Discriminant plots produced by FDA.	160
7.17	Coefficients from the DWT which were selected by SWBLDA (asterisk) and SWBQDA (circle).	162
7.18	Reconstructed spectra produced from the coefficients selected by SWBLDA and SWBQDA.	163
7.19	The wavelet coefficients and reconstructed spectra produced from the AWA.	165
7.20	Discriminant plots produced by from the coefficients produced by the AWA.	166
7.21	Discriminant plots produced by PDA.	167
8.1	Five sample spectra from the sugar data.	17
8.2	Five sample spectra from the wheat data.	172
8.3	Test r-squared values corresponding to the brix, fibre and protein responses.	175
8.4	The DWT and inverse DWT performed on the sugar data.	179
8.5	The DWT and inverse DWT performed on the wheat data.	180
8.6	Coefficients selected from the DWT by SMLRW.	183

8.7	Regression criterion measure versus iteration for the adaptive wavelet algorithm.	187
8.8	Test r-squared values for the wavelet based regression methods.	189
8.9	Test r-squared values each of the regression strategies.	189
8.10	Test r-squared values each of the regression strategies (SMLRW and SPCRW not shown).	190
8.11	Residuals versus fitted values for the brix response models.	194
8.12	Residuals versus fitted values for the fibre response models.	195
8.13	Residuals versus fitted values for the protein response models.	196
8.14	Histograms of the residuals from the brix response models.	197
8.15	Histograms of the residuals from the fibre response models.	198
8.16	Histograms of the residuals from the protein response models.	199
8.17	Plots of the residuals versus the fitted values for each of the models for brix.	200
8.18	Plots of the residuals versus the fitted values for each of the models for fibre.	201
8.19	Plots of the residuals versus the fitted values for each of the models for protein.	202
8.20	Wavelengths selected by SMLR-S1 and SMLR-S2.	203
8.22	Regression coefficients obtained from PLS, when the data has been standardized.	203
8.21	Absolute correlations between each wavelength and the principal components selected by SPCR.	204
8.24	Reconstructed spectra produced from the coefficients selected by SMRLW.	205
8.23	Coefficients from the DWT which were selected by SMLRW.	206
8.25	Reconstructed spectra produced from the coefficients selected by SMRLW that pertain to the same band.	208

8.26 The wavelet coefficients and reconstructed spectra produced from the AWA.	210
---	-----

List of Tables

5.1	Summary statistics for the correlation coefficients of the scaling and wavelet coefficients of a spectral data set.	100
6.1	The percentage of correctly classified spectra, using the coefficients $\{X^{[3]}(\tau)\}$ for $\tau = 0, \dots, 3$ at initialization and at termination of the adaptive wavelet algorithm. The discriminant criterion functions were Wilk's Lambda, symmetric entropy and the CVQPM.	119
6.2	The percentage of correctly classified spectra, using the coefficients $\{X^{[3]}(\tau)\}$ for $\tau = 0, \dots, 3$ at initialization and at termination of the adaptive wavelet algorithm. Optimization was based on $\{X^{[3]}(0)\}$ and the discriminant criterion functions were Wilk's Lambda, symmetric entropy and the CVQPM.	120
7.1	Description of the spectral data sets used for classification.	122
7.2	Correct classification rates (%) for the stepwise procedures.	128
7.3	Original variables selected by SBLDA and SBQDA.	129
7.4	Correct classification rates (%)	130
7.5	Quadratic probability measures	130
7.6	Classification results for BBLDA.	138
7.7	Classification results for BBQDA.	139
7.8	Correct classification rates for SWBLDA and SWBQDA.	140
7.9	Coefficients selected by the forward schemes for SWBLDA and SWBQDA.	140

7.10	Classification performance of the LDB algorithm.	145
7.11	Classification results for the adaptive wavelet algorithm.	148
7.12	Classification results for the adaptive wavelet algorithm where optimization was over a scaling and wavelet band for the (4,3,2) setting.	150
7.13	Correct classification rates for the wavelet based feature extraction strategies.	151
7.14	Quadratic probability measures for the wavelet based feature extraction strategies.	152
8.1	Description of the spectral data sets used for regression.	171
8.2	Training and test R-squared values.	175
8.3	Wavelengths selected by the SMLR routines, and the principal components selected by SPCR.	177
8.4	Classification results for banded BLDA.	181
8.5	R-squared values for SMLRW-S1 and SMLRW-S2.	182
8.6	Coefficients selected from the DWT by SMLRW-S1 and SMLRW-S2.	182
8.7	R-squared values for SPCRW-S1 and SPCRW-S2.	184
8.8	Components selected from the DWT by SPCRW-S1 and SPCRW-S2.	185
8.9	Regression results for the adaptive wavelet algorithm.	186
8.10	Training and test r-squared values for the wavelet based regression approaches.	188
8.11	Summary of p-values for the regression models.	192

List of Symbols

Non-bold Lower Case Letters

- $a(\mathbf{x})$ appreciation score of \mathbf{x}
- $a_{\text{ccr}}(\mathbf{x})$ appreciation score equal to 1 if $P(r | \mathbf{x}_{i(r)}) \geq P(r | \mathbf{x}_i)$ and zero otherwise
- $a_A(\mathbf{x})$ appreciation score equivalent to $P(r | \mathbf{x}_{i(r)})$
- $a_Q(\mathbf{x})$ quadratic appreciation score of \mathbf{x}
- a_{il} l th element in the i th principal component vector
- a parameter used in RDA which weights the pooled covariance matrix
- b parameter used in RDA which controls shrinkage of the weighted pooled covariance matrix
- b_i i th element in the vector of estimated regression coefficients \mathbf{b}
- $\text{band}(j, t)$ τ th band $\tau \in \{0, 1, \dots, m - 1\}$ at the j th level $j \in \{J, J - 1, \dots, J - \max_{\text{lev}} + 1\}$ of the DWT
- $c_{j,k}$ scaling coefficients
- $d_{j,k}$ wavelet coefficients
- f_{V_j} orthogonal projection of $f(t)$ onto V_j
- $g(\mathbf{x}, r)$ classification score
- $g_{\text{blda}}(\mathbf{x}, r)$ BLDA classification score
- $g_{\text{bqda}}(\mathbf{x}, r)$ BQDA classification score
- h_k high pass filter coefficients
- \tilde{h}_{ii} is the element along the i th diagonal of the hat matrix \mathcal{H}
- j_* complex number $\sqrt{-1}$

- j parameter controlling the dilation of the wavelet basis functions
- k parameter controlling the translation of the wavelet basis functions
- ℓ_k low pass filter coefficients
- m number of bands in the DWT; downsampling rate
- \max_{lev} maximum number of levels in the DWT.
- n number of observational units in the training data set
- n' number of observational units in the testing data set
- n_r number of observational units from class r ; r th element in the vector \mathbf{n}
- $n_{[l]}$ number of objects in node l of CART model
- n_{levels} number of levels that an object has been transformed, in the DWT
- p dimensionality of the data set
- p_* dimensionality of the reduced data set $p_* \ll p$
- p_o number of parameters to be estimated (including the intercept) in a MLR model
- $p(\mathbf{x})$ is the class probability density of \mathbf{x}
- q the number of sub-matrices in the filter coefficient matrix \mathbf{A} is $q + 1$
- r index for class categories
- s_o minimum of one less than the total number of classes $(R-1)$, or the dimensionality (p) .
- s_* number of discriminant variables used for assigning an object to a class; $s_* \leq s_o$
- x_i i th element in the data vector \mathbf{x}
- $x_{i[l]}$ i th object in node l of CART model
- y_i i th element in the response vector \mathbf{y}
- y'_i i th element in the test response vector \mathbf{y}'

- $y_{i[l]}$ response value of i th object in node l of CART model
- \hat{y}_{-i} predicted value of \mathbf{x}_i , obtained when \mathbf{x}_i is deleted from the model building process
- \hat{y}_i predicted response value for object \mathbf{x}_i
- \hat{y}'_i predicted response value for object \mathbf{x}'_i
- y_{ij} element in row i and column j of \mathbf{Y}
- \hat{y}_{ij} estimate of y_{ij}
- z index for wavelet filter $z = 1, \dots, m - 1$

Non-bold Upper Case Letters

- AIC Akaike's information criterion
- CCR correct classification rate
- CCR' correct classification rate of test set
- C_p Mallows C_p
- CVCCR cross-validated correct classification rate
- DF degrees of freedom
- DEV deviation
- $\mathcal{D}(\mathbf{x}, r)$ distance between \mathbf{x} and $\bar{\mathbf{x}}_r$ in the discriminant coordinate system
- E_{cross} cross entropy measure
- E_{sym} symmetric entropy measure
- E_{tot} total symmetric entropy measure
- \mathcal{F}_{CWT} continuous wavelet transform
- \mathcal{F}_{DWT} discrete wavelet transform
- \mathcal{F}_{FT} Fourier transform

- \mathcal{F}_{WFT} windowed Fourier transform
- J highest level in the DWT ; $J = \text{ceiling}(\log p / \log m)$
- \mathcal{J} criterion function applied in the adaptive wavelet or LDB algorithm
- \mathcal{J}_Λ Wilk's lambda discriminatory criterion function
- \mathcal{J}_E entropy discriminatory criterion function
- $\mathcal{J}_{\text{cvqpm}}$ discriminatory criterion function based on the cross-validated quadratic probability measure
- $\mathcal{J}_{\text{cvrsq}}$ regression criterion function based on the cross-validated r-squared measure
- $L^2(\mathbb{R})$ space of square integrable functions
- M_{ij} i, j th element in the Lawton matrix
- MCR misclassification rate
- MSE mean square error
- N_i node identity in CART model
- N_f number of filter coefficients with nonnegative indices
- P_A average probability that an object is assigned to the correct class
- P_{QPM} quadratic probability measure
- P_{CCR} probability of correctly classifying objects
- $P(r)$ prior probability for class r
- $P(r | \mathbf{x})$ posterior probability that given some vector \mathbf{x} it is from class r
- $P(r | \mathbf{x}_{i(r)})$ posterior probability for the true class of \mathbf{x}_i
- $P(\mathbf{x} | r)$ class probability density function
- $P(r | l)$ proportion of objects in node N_l of a CART model which are from class r

- $P_{-i}(r \mid \mathbf{x}_i)$ posterior probability for \mathbf{x}_i when the covariance matrices and mean vectors in the probability density function have been calculated in the absence of \mathbf{x}_i
- PRESS predicted residual sum of squares
- RSS residual sum of squares
- RSS_{p_o} residual sum of squares of a MLR model with complexity p_o
- R^2 coefficient of variation (r-squared)
- R total number of class categories in a set of data
- R^* integer value less than or equal to $R - 1$
- TSS total sum of squares
- V number of testing groups used in a cross-validation routine
- V_j subspace containing all the possible approximations of functions in $L^2(\mathbb{R})$ at resolution 2^j
- W_j orthogonal complement of v_j

Bold Lower Case Letters

- \mathbf{a}_i i th vector of principal component coefficients with dimension $p \times 1$
- \mathbf{b} estimated vector of regression coefficients
- $\mathbf{b}_{r_{os}}$ r th column of the matrix of regression coefficients for the optimal scoring problem, \mathbf{B}_{os}
- \mathbf{b}_{pls} estimated vector of regression coefficients from a PLS model
- \mathbf{c}_j scaling coefficients at resolution (or level) j
- \mathbf{d}_j wavelet coefficients at resolution (or level) j
- $\mathbf{d}_j^{(z)}$ wavelet coefficients at resolution (or level) j produced from the filter matrix $\mathbf{D}_{j+1}^{(z)}$

- $\mathbf{e}_{(r)}^{[j]}(\tau)$ class energy vector of wavelet (or wavelet packet) coefficients
- ℓ vector of low pass filter coefficients
- n $R \times 1$ vector of class sample sizes
- \mathbf{p} $n \times 1$ vector containing principal component scores
- \mathbf{r} vector of residuals in the PLS algorithm
- \mathbf{s} output from low pass filtering operation
- t latent variables from PLS model
- \mathbf{u}_i normalized vectors which are used to construct the wavelet matrix \mathbf{A}
- \mathbf{v} normalized vector which is used to construct the wavelet matrix \mathbf{A}
- \mathbf{v} $p \times 1$ vector of discriminant coefficients
- \mathbf{w}_1 sums of squares and cross product between \mathbf{X} and \mathbf{y}
- \mathbf{w} output from high pass filtering operation
- \mathbf{x} $p \times 1$ training data vector
- $\bar{\mathbf{x}}$ $p \times 1$ mean vector of the training data set
- \mathbf{x}' $p \times 1$ testing data vector
- \mathbf{x}^* $p \times 1$ column object vector from \mathbf{X}^*
- $\mathbf{x}^{[j]}(\tau)$ column vector containing the coefficients in $\text{band}(j, \tau)$ of the DWT
- $\mathbf{x}_{i(r)}$ $p \times 1$ data vector from class r
- $\mathbf{x}_{i(r)}^*$ i th data object from \mathbf{X}^* which belongs to class r
- $\bar{\mathbf{x}}_r^*$ mean class vector from \mathbf{X}^*
- ${}^o\mathbf{x}^{[j]}(\tau)$ wavelet packet coefficients which occur at the j th level in the τ th band of the wavelet packet transform

- \mathbf{y} $n \times 1$ vector of training response values (regression) or class labels (discriminant analysis)
- $\hat{\mathbf{y}}$ $n \times 1$ predicted vector of response values (regression) or class labels (discriminant analysis)
- \mathbf{y}' $n' \times 1$ vector of test response values (regression) or class labels (discriminant analysis)
- $\hat{\mathbf{y}}'$ $n' \times 1$ predicted vector of test response values (regression) or class labels (discriminant analysis)
- \mathbf{z} $n \times 1$ discriminant variable

Bold Upper Case Letters

- \mathbf{A} wavelet matrix
- \mathbf{A}_i sub-matrix of the wavelet matrix \mathbf{A}
- \mathbf{B} matrix of multivariate regression coefficients
- \mathbf{B}_{os} optimal scoring matrix of regression coefficients
- \mathbf{C}_j low pass filtering matrix at level j in the DWT
- \mathbf{D}_j high pass filtering matrix at level j in the DWT
- $\mathbf{D}_j^{(z)}$ high pass filtering matrix at level j in the DWT which contains the z th set of highpass filter coefficients
- \mathbf{D} diagonal matrix whose i th diagonal element is equal to $D_{ii} = 1/\sqrt{\lambda_{ifda}^2(1 - \lambda_{ifda}^2)}$
- \mathbf{F}_i i th factor in the wavelet matrix \mathbf{A}
- \mathbf{L} low pass convolution matrix
- \mathcal{H} hat matrix $\mathcal{H} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}$
- \mathbf{H} high pass convolution matrix

- \mathbf{P} matrix whose i th column contains the principal component scores vector \mathbf{p}_i
- \mathbf{P}_1 is a matrix which augments $\mathbf{1}_n$ to the first column of \mathbf{P}
- $\mathbf{P}_X, \mathbf{P}_{X^*}$ linear projector matrices
- \mathbf{Q} orthogonal matrix used in construction of the wavelet matrix \mathbf{A}
- \mathbf{R} projection matrix used in construction of the wavelet matrix \mathbf{A}
- \mathbf{S}_B between covariance matrix
- \mathbf{S}_W within covariance matrix
- $\mathbf{S}_{\text{pooled}}$ pooled covariance matrix
- \mathbf{S}_r covariance matrix of class r
- \mathbf{T} matrix whose i th column contains the i th latent vector from PLS
- \mathbf{V}_{s_o} matrix whose i th column is \mathbf{v}_i for $i = 1, \dots, s_o$.
- \mathbf{X} $p \times n$ training data matrix
- \mathbf{X}_1 training data matrix whose first row is equal to $\mathbf{1}_n^T$
- \mathbf{X}_c $p \times n$ centered training data matrix
- \mathbf{X}' $p \times n'$ testing data matrix
- \mathbf{X}^* $p \times n$ data matrix which results from some feature selection/transformation procedure based on \mathbf{X} .
- $\mathbf{X}^{[j]}(\tau)$ matrix containing the coefficients for the objects which would lie in band(j, τ)
- \mathbf{Y} $n \times R$ class indicator matrix
- \mathbf{Z}_{s_o} matrix whose i th column is \mathbf{z}_i for $i = 1, \dots, s_o$

Greek Letters

- β_i i th component in the vector of regression coefficients β
- $\delta(t)$ delta function
- δ_{ij} indicator variable; $\delta_{ij} = 1$ if $i = j$, zero otherwise
- ϵ_i i th component in the vector of regression residuals ϵ
- γ_i eigenvalue corresponding to the i th principal component
- λ is a measure of the discriminant criterion $\lambda = \mathbf{v}^T \mathbf{S}_B \mathbf{v}$
- $\lambda_{i_{\text{fda}}}$ i th element of λ_{fda}
- Λ Wilk's Lambda
- $\Lambda^{(i)}$ Wilk's Lambda at the i th iteration of a stepwise routine
- $\mathcal{U}(j, \tau)$ discriminatory measure of band(j, τ) in the wavelet packet transform
- ν_i i th element in ν
- ω frequency
- $\phi(t)$ scaling function
- $\phi_{j,k}(t)$ scaling basis function; $\phi_{j,k}(t) = m^{j/2} \phi(m^j t - k)$
- $\psi(t)$ mother wavelet function
- $\psi_{j,k}(t)$ wavelet basis function; children wavelets; $\psi_{j,k}(t) = m^{j/2} \psi(m^j t - k)$
- $\hat{\rho}_{ij}$ correlation between the i th principal component and the j th variable
- $\hat{\sigma}_{x_i}$ sample standard deviation of x_i
- τ band label for the DWT; $\tau \in 0, 1, \dots, m - 1$
- ρ rank of a matrix
- β vector of regression coefficients

- β_{pcr} vector of regression coefficients from a PCR model
- β_{pls} vector of regression coefficients from a PLS model
- λ_{fda} vector whose elements are the eigenvalues of $\Psi^{*T}\Psi^*/n$
- Λ_{fda} diagonal matrix whose i th element is equal to $\lambda_{i_{\text{fda}}}$
- $\eta(x^*)$ vector of fitted values for x^*
- $\bar{\eta}_r$ fitted centroid of all x^* objects belonging to class r
- ν vector of wavelengths
- Ψ^* class indicator matrix used in FDA and PDA
- $\hat{\Psi}^*$ estimate of the class indicator matrix Ψ^*
- Θ matrix whose columns are the eigenvectors of $\Psi^{*T}\Psi^*/n$

Miscellaneous Characters

- $\mathbf{1}_i$ $i \times 1$ column vector whose elements are all equal to 1
- $\downarrow m$ downsample by a factor of m

List of Algorithms

Flexible Discriminant Algorithm.....	20
Partial Least Squares Algorithm.....	42
Second Derivative Algorithm.....	61
Best Basis Algorithm.....	106
Adaptive Wavelet Algorithm.....	117