



Universiteit
Leiden
The Netherlands

Feature network models for proximity data : statistical inference, model selection, network representations and links with related models

Frank, L.E.

Citation

Frank, L. E. (2006, September 21). *Feature network models for proximity data : statistical inference, model selection, network representations and links with related models*. Retrieved from <https://hdl.handle.net/1887/4560>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4560>

Note: To cite this publication please use the final published version (if applicable).

Feature Network Models for Proximity Data

Frank, Laurence Emmanuelle,
Feature Network Models for Proximity Data. Statistical inference, model selection,
network representations and links with related models.
Dissertation Leiden University - With ref. - With summary in Dutch.
Subject headings: additive tree; city-block models; distinctive features models; fea-
ture models; feature network models; feature selection; Monte Carlo simulation;
statistical inference under inequality constraints.

ISBN 90-8559-179-1

© 2006, Laurence E. Frank

Printed by Optima, Rotterdam

Manuscript prepared in L^AT_EX (pdf_{tex}) with the T_EX previewer TeXShop (v1.40), using the memoir document class (developed by P. Wilson) and the apacite package for APA style bibliography (developed by E. Meijer).

Feature Network Models for Proximity Data

*Statistical inference, model selection, network representations
and links with related models*

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 21 september 2006
klokke 13.45 uur

door

Laurence Emmanuelle Frank

geboren te Delft
in 1969

PROMOTIECOMMISSIE

Promotor: Prof. Dr. W. J. Heiser

Referent: Prof. J. E. Corter, Ph.D.,
Columbia University, New York, USA

Overige leden: Prof. Dr. I. Van Mechelen, K.U. Leuven, België
Prof. Dr. V. J. J. P. van Heuven
Prof. Dr. J. J. Meulman
Prof. Dr. P. M. Kroonenberg

To my parents

“On ne peut se flatter d’avoir le dernier mot d’une théorie, tant qu’on ne peut pas l’expliquer en peu de paroles à un passant dans la rue.”

[It is not possible to feel satisfied at having said the last word about some theory as long as it cannot be explained in a few words to any passer-by encountered in the street.]

Joseph Diaz Gergonne, French mathematician (Chasles, 1875, p. 115).

Acknowledgements

A large number of persons contributed in several ways to this dissertation and I am indebted to them for their support.

I learned a lot about research in psychometrics from being a member of the Interuniversity Graduate School for Psychometrics and Sociometrics (IOPS). I would like to thank the IOPS-students for the agreeable time and a special thank to Marieke Timmerman, for her interest and the pleasant conversations. A special thank also to Susañna Verdel for the enjoyable way we prepared the IOPS meetings, and for the wonderful way she organizes all practical IOPS-issues, always trying to offer the best possible conditions for staff and students.

I am very grateful for the opportunities given to me to attend conferences of the Psychometric Society and the International Federation of Classification Societies, which introduced me to the scientific community of our field of research and has been very inspiring for my own research.

I am greatly indebted to Prof. L.J. Hubert, Ph.D. (University of Illinois at Urbana-Champaign, USA) for helping me to implement the Dykstra algorithm in *Matlab* during his stay in Leiden, and for his useful comments on earlier versions of the second and third chapter of this dissertation.

For support on a more daily basis I would like to thank my colleagues of Psychometrics and Research Methodology and Data Theory Group (Universiteit Leiden): Bart Jan van Os for helping me with technical issues at crucial points during this research project, but also for his interest and the pleasant coffee breaks; Mark de Rooij for showing me how to do research in our field and how to accomplish a Ph.D. project; Mariëlle Linting for sharing a lot of nice conference experiences and hotel rooms during the whole project; Matthijs Warrens for the inspiring conversations about research; Marike Polak for the daily pleasant, encouraging conversations with lots of coffee and tea, and her sincere interest, which also holds for Rien van der Leeden.

The accomplishment of this dissertation would not have been possible without the love and support of my family and friends. To all who supported me during these years: many thanks for your friendship and all the joyous moments shared. It helped me to place this work in the right perspective.

Contents

Acknowledgements	ix
Contents	xi
List of Figures	xv
List of Tables	xix
Notation and Symbols	xxi
Notation conventions	xxi
Symbols	xxi
1 Introducing Feature Network Models	1
1.1 Features	2
Distinctive features versus common features	3
Where do features come from?	6
Feature distance and feature discriminability	7
1.2 Feature Network	8
Parsimonious feature graphs	8
Embedding in low-dimensional space	10
Feature structure and related graphical representation	12
Feature networks and the city-block model	13
1.3 Feature Network Models: estimation and inference	14
Statistical inference	14
Finding predictive subsets of features	17
1.4 Outline of the monograph	19
2 Estimating Standard Errors in Feature Network Models	21
2.1 Introduction	21
2.2 Feature Network Models	23
2.3 Obtaining standard errors in Feature Network Models with a priori features	28
Estimating standard errors in inequality constrained least squares	28
Determining the standard errors by the bootstrap	30
Bootstrap procedures	32

	Results bootstrap	32
2.4	Monte Carlo simulation	38
	Sampling dissimilarities from the binomial distribution	38
	Simulation procedures	40
	Additional simulation studies	41
2.5	Results simulation	44
	Bias	44
	Coverage	45
	Power and alpha	48
2.6	Discussion	48
3	Standard Errors, Prediction Error and Model Tests in Additive Trees	51
3.1	Introduction	51
3.2	Feature Network Models	54
3.3	Feature Network Models: network and additive tree representations	57
	Additive tree representation and feature distance	59
3.4	Statistical inference in additive trees	63
	Obtaining standard errors for additive trees	63
	Testing the appropriateness of imposing constraints	65
	Estimating prediction error	66
3.5	Method Monte Carlo simulations	67
	Empirical p -value Kuhn-Tucker test	68
	Simulation for nominal standard errors with a priori tree topology	68
	Simulation for nominal standard errors with unknown tree topology	70
3.6	Results simulation	74
	Results Kuhn-Tucker test and estimates of prediction error	74
	Performance of the nominal standard errors for known tree topology	74
	Performance of the nominal standard errors for unknown tree	
	topology	76
3.7	Discussion	79
4	Feature Selection in Feature Network Models: Finding Predictive Sub-	
	sets of Features with the Positive Lasso	83
4.1	Introduction	83
4.2	Theory	86
	Feature Network Models	86
	Generating features with Gray codes	90
	Selecting a subset of features with the Positive Lasso	93
	Generating features by taking a random sample combined with a	
	filter	99
	Example of feature generation and selection on the <i>consonant</i> data	99
4.3	Simulation study	101
	Method for simulation study	102
	Results simulation study	105
4.4	Discussion	111

5	Network Representations of City-Block Models	115
5.1	Network representations of city-block models	115
5.2	General theory	118
	Betweenness of points and additivity of distances	118
	Network representation of city-block configurations	119
	Internal nodes	123
	Partial isometries	127
5.3	Discrete models that are special cases of the city-block model . . .	127
	Lattice betweenness of feature sets	128
	Distinctive features model	129
	Additive clustering or the common features model	133
	Exact fit of feature models	138
	Partitioning in clusters with unicities: the double star tree	140
	Additive tree model	141
5.4	Discussion	143
6	Epilogue: General Conclusion and Discussion	149
6.1	Reviewing statistical inference in Feature Network Models	149
	Constrained estimation	149
	Bootstrap standard deviation	151
	Assumptions and limitations	152
6.2	Features and graphical representation	153
	The set of distinctive features	153
	FNM and tree representations	156
	References	159
	Author Index	171
	Subject Index	175
	Summary in Dutch (Samenvatting)	179
	Curriculum vitae	185

List of Figures

1.1	Feature network of all presidents of the USA based on 14 features from Schott (2003, pp. 14-15). The presidents are represented as vertices (black dots) and labeled with their names and chronological number. The features are represented as internal nodes (white dots).	1
1.2	Experimental conditions <i>plants</i> data. The 16 plants vary in the form of the pot and in elongation of the leaves. (Adapted with permission from: Tversky and Gati (1982), Similarity, separability, and the triangle inequality. <i>Psychological Review</i> , 89, 123-154, published by APA.)	4
1.3	Complete network <i>plants</i> data.	9
1.4	Triangle equality and betweenness.	10
1.5	Feature graph of the <i>plants</i> data using the features resulting from the experimental design with varying elongation of leaves and form of the pot (with 6 of the 8 features).	11
1.6	Additive tree representation of the <i>plants</i> data.	12
1.7	Feature network representing a 6-dimensional hypercube based on the unweighted, reduced set of features of the <i>plants</i> data. Embedding in 2-dimensional Euclidean space was achieved with PROXSCAL allowing ordinal proximity transformation with ties untied and the Torgerson start option.	13
1.8	An overview of the steps necessary to fit Feature Network Models with PROXGRAPH.	15
1.9	Feature graph for the <i>plants</i> data, resulting from the Positive Lasso feature subset selection algorithm on the complete set of distinctive features. The original experimental design is the cross classification of the form of the pot (a,b,c,d) and the elongation of the leaves (p,q,r,s). Embedding in 2-dimensional space was done with PROXSCAL using ratio transformation and the simplex start option. ($R^2 = 0.81$)	18
2.1	Feature Network Model on consonant data ($dh = \bar{\theta}$; $zh = \bar{\zeta}$; $th = \theta$; $sh = \bar{\jmath}$). . .	27
2.2	Empirical distribution of OLS (top) and ICLS (bottom) estimators (1,000 bootstrap samples).	35
2.3	Comparison of nominal confidence intervals for ICLS estimator with bootstrap- t CI (top) and bootstrap BC_a CI (bottom); long bar = nominal CI; short bar = bootstrap- t CI or BC_a CI.	36

2.4	BC_a and nominal confidence intervals for OLS and ICLS estimators (long bar = nominal CI; short bar = BC_a CI).	37
2.5	Sampling dissimilarities from a binomial distribution	40
2.6	Coverage Nominal CI, Bootstrap- t CI, and BC_a CI for ICLS estimates for all simulation studies. The order of the plots follows the increasing number of zero and close to zero parameters present in the data.	47
3.1	Feature Network representation for the <i>kinship</i> data with the three most important features (<i>Gender</i> , <i>Nuclear family</i> and, <i>Collaterals</i>) represented as vectors. The plus and minus signs designate the projection onto the vector of the centroids of the objects that possess the feature (+) and the objects that do not have that feature (-).	57
3.2	Nested and disjoint feature structure and corresponding additive tree representation. Each edge in the tree is represented by a feature and the associated feature discriminability parameter η_t	58
3.3	Betweenness holds when $J = I \cap K$, where I , J , and K are sets of features describing the corresponding objects i , j , and k	59
3.4	Unresolved additive tree representation of the <i>kinship</i> data based on the solution obtained by De Soete & Carroll (1996).	61
3.5	Feature structure for the resolved additive tree representation (<i>top</i>) of the <i>kinship</i> data and simplified feature structure for the unresolved additive tree representation (<i>bottom</i>) of Figure 3.4.	62
3.6	Feature parameters ($\hat{\eta}_{\text{ICLS}}$) and 95% t -confidence intervals for additive tree solution on <i>kinship</i> data with $R^2 = .96$	63
3.7	Additive tree representation of the <i>fruit</i> data obtained with PROXGRAPH based on the tree topology resulting from the neighbor-joining algorithm.	70
3.8	Histogram of Kuhn-Tucker test statistic obtained with parametric bootstrap (1,000 samples) with ICLS as H_0 model, based on <i>kinship</i> data. The empirical p -value is equal to .74 and represents the proportion of samples with values on the Kuhn-Tucker statistic larger than 0.89, the value of the statistic observed for the sample.	74
3.9	Mean (panel A), bias (panel B), and $rmse$ (panel C) of the 1,000 simulated nominal standard errors $\hat{\sigma}_{\text{ICLS}}$ (\bullet) and the 1,000 bootstrap standard deviations sd_B (\square) plotted against the true nominal standard errors σ_{ICLS}	75
3.10	Coverage proportions of the nominal t -CI and bootstrap t -CI for the true feature discriminability values, based on the 1,000 simulated samples.	76
3.11	<i>Left panel</i> : Distribution of the GCV_{FNM} statistic estimated on the test samples based on the tree topology inferred for the training samples under all experimental conditions for 100 simulation samples. The asterisk in each box represents the mean of the true GCV_{FNM} values. <i>Right panel</i> : Distribution of the number of cluster features equal to the true cluster features ($T_C = 17$) present in the tree topologies obtained for the training samples of the same 100 simulation samples in each experimental condition.	77

3.12	Coverage proportions in all experimental conditions for feature discriminability parameters based on nominal t -CI (\bullet) in the test samples and proportions recovered true features in the training samples (\square) for each of the 37 features forming the true tree topology.	82
4.1	Feature Network representation for the <i>consonant</i> data with the three most important features (<i>voicing, nasality, and duration</i>) represented as vectors. The plus and minus signs designate the projections onto the vector of the centroids of the objects that possess the feature (+) and the objects that do not have that feature (-). (dh = δ ; zh = ζ ; th = θ ; sh = \jmath).	90
4.2	Graphs of estimation for the Lasso (left) and ridge regression (right) with contours of the least squares error functions (the ellipses) and the constraint regions, the diamond for the Lasso and the disk for ridge regression. The corresponding constraint functions are equal to $ \beta_1 + \beta_2 \leq b$ for the Lasso and $\beta_1^2 + \beta_2^2 \leq b^2$ for ridge regression. It is clear that only the constraint function of the Lasso can force the $\hat{\beta}$ -values to become exactly equal to 0. (The graphs are adapted from Hastie et al. (2001), p. 71).	95
4.3	Estimates of feature parameters for the <i>consonant data</i> . <i>Top panels</i> : trajectories of the Lasso estimates $\hat{\eta}_L$ (left panel) and the AIC_L values plotted against the effective number of parameters (= df) of the Lasso algorithm (right panel). The model with lowest AIC_L value (= 0.65) contains all 7 features. <i>Lower panels</i> : trajectories of the Positive Lasso estimates $\hat{\eta}_{PL}$ (left panel) and the adjusted AIC_L values plotted against the effective number of parameters (= df) of the Positive Lasso algorithm (right panel). The model with lowest AIC_L value (= 0.71) has 5 features.	97
4.4	AIC_L -plot for the <i>consonant</i> data using all possible features generated with Gray codes ($T = 32,767$). The lowest AIC_L value (= 0.51) points to a model with 7 features.	100
4.5	Feature Network representation for the <i>consonant</i> data based on the feature matrix selected by the Positive Lasso displayed in Table 4.6. (dh = δ ; zh = ζ ; th = θ ; sh = \jmath).	101
4.6	Feature network plots for the experimental conditions for 12 objects. A = 4 features, medium η ; B = 4 features, small + large η ; C = 8 features, medium η ; D = 8 features, small + large η	105
4.7	Boxplots showing the distributions of 50 simulation samples on 12 objects using the complete set of Gray codes. The experimental conditions are medium (left panels) and small + large (right panels) η values, two error conditions, low (L) and high (H), and two levels of true number of features (4 and 8) corresponding to two levels of n/T ratio equal to 16 and 8. The top panels show the effective number of features selected for each sample (= Df) with the true number of features represented as a dashed line. The lower panels show the associated AIC_L values.	107

4.8	Boxplots showing the distributions of 50 simulation samples on 12 objects using a large random sample of the complete set of Gray codes combined with a filter. The experimental conditions are medium (left panels) and small + large (right panels) η values, two error conditions, low (L) and high (H), and two levels of true number of features (4 and 8) corresponding to two levels of n/T ratio equal to 16 and 8. The top panels show the effective number of features selected for each sample ($= Df$) with the true number of features represented as a dashed line. The lower panels show the associated AIC_L values.	109
4.9	Boxplots showing the distributions of 50 simulation samples on 24 objects using a large random sample of the complete set of Gray codes. The experimental conditions are medium (left panels) and small + large (right panels) η values, two error conditions, low (L) and high (H), and two levels of true number of features (17 and 35) corresponding to two levels of n/T ratio equal to 16 and 8. The top panels show the effective number of features selected for each sample ($= Df$) with the true number of features represented as a dashed line. The lower panels show the associated AIC_L values.	110
5.1	City-block solution in two dimensions for the <i>rectangle</i> data. The labels $W_1 - W_4$ indicate the width levels, and $H_1 - H_4$ the height levels of the stimulus rectangles.	121
5.2	Equal city-block distances among four points. Tetrahedron with equal edge lengths (<i>left panel</i>) and star graph with equal spokes, which generates the same distances (<i>right panel</i>).	124
5.3	Network representation of the two-dimensional city-block solution for the <i>rectangle</i> data, including fifteen internal nodes. The labels $W_1 - W_4$ indicate the width levels, and $H_1 - H_4$ the height levels of the stimulus rectangles.	125
5.4	Partial isometry: two different configurations with the same city-block distances. <i>Left panel</i> : Network representation of A, B, C and the points P1–P5. <i>Right panel</i> : Network representation of A, B, C and the points P1–P5. The two networks share the internal point H, the hub.	126
5.5	Network representation of distinctive features model for the <i>number</i> data, without internal nodes. Nodes labeled by stimulus value.	131
5.6	Network representation of distinctive features model for the <i>number</i> data, with internal nodes. Solid dots are stimuli labeled by stimulus value, open dots are internal nodes labeled by subset.	134
5.7	Network representation of common features model for <i>body-parts</i> data, with internal nodes.	137
5.8	Network representation of double star tree for the <i>number</i> data.	141
5.9	Network representation of additive tree for the <i>number</i> data.	144
5.10	Relationships between city-block models.	146
6.1	Biplot in 2 dimensions obtained with correspondence analysis of the 14 features describing the 43 presidents of the United States. The presidents are linked with the features they possess. (Normalization: row principal).	157

List of Tables

1.1	Feature matrix of 16 plants (Figure 1.2) varying in form of the pot (features: a, b, c) and elongation of the leaves (features: p, q, r), see Tversky and Gati (1982).	3
1.2	Overview of graphical and non-graphical models based on common features (CF) and distinctive features (DF)	5
1.3	Feature discriminability estimates, standard errors and 95% confidence intervals for <i>plants</i> data using six features selected from the complete experimental design in Table 1.1 and associated with the network graph in Figure 1.5 ($R^2 = 0.60$).	16
1.4	Feature matrix resulting from feature subset selection with the Positive Lasso on the <i>plants</i> data.	17
2.1	Matrix of 16 English consonants, their pronunciation and phonetic features .	24
2.2	Feature parameters, standard errors and 95% confidence intervals for consonant data	26
2.3	Three types of 95% Confidence Intervals for ICLS and OLS estimators resulting from the bootstrap study on the <i>consonant</i> data.	33
2.4	Description of features and the corresponding objects for three additional data sets	43
2.5	Bias and <i>rmse</i> of $\hat{\eta}$, $\hat{\sigma}_{\hat{\eta}}$, and bootstrap standard deviation (sd_B) for OLS and ICLS estimators, resulting from the Monte Carlo simulation based on the <i>consonant</i> data.	44
2.6	Coverage , empirical power and alpha for nominal and empirical 95% confidence intervals (Monte Carlo simulation based on <i>consonant</i> data)	46
3.1	The 5 binary features describing the kinship terms	55
3.2	Feature parameters ($\hat{\eta}$), standard errors and 95% <i>t</i> -confidence intervals for Feature Network Model on <i>kinship</i> data with $R^2 = .95$	56
3.3	The 17 cluster features ($F_1 - F_{17}$) and 20 unique features ($F_{18} - F_{37}$) with associated feature discriminability parameters for the neighbor-joining tree on the <i>fruit</i> data.	73
3.4	Proportion of 95% <i>t</i> -confidence intervals containing the value zero in the test samples for the feature discriminability parameters associated with features not present in the true tree topology	78

4.1	Matrix of 16 English consonants, their pronunciation and phonetic features	86
4.2	Feature parameters ($\hat{\eta}$), standard errors, and 95% confidence intervals for Feature Network Model on <i>consonant</i> data with $R^2 = 0.61$	89
4.3	Binary code and Gray code for 4 bits	91
4.4	Estimates of feature discriminability parameters ($\hat{\eta}_{\text{ICLS}} = \text{ICLS}$, $\hat{\eta}_{\text{L}} = \text{Lasso}$, and $\hat{\eta}_{\text{PL}} = \text{Positive Lasso}$) for the <i>consonant data</i>	98
4.5	Positive Lasso estimates, R^2 , and prediction error (K -fold cross-validation) for the features from phonetic theory (left) and for the features selected from the complete set of distinctive features (right)	102
4.6	Matrices of features based on phonetic theory (left) and of features selected by the Positive Lasso (right)	103
4.7	Feature matrices for 12 objects and rank numbers used to construct the true configurations for the simulation study	104
4.8	Proportion of correctly recovered features from the complete set of distinctive features under combined levels of error (L = low; H = high), the ratio of the number of object pairs and the number of features (= n/T ratio), and feature parameter (η) sizes, medium and small + large.	106

Notation and Symbols

Notation conventions

matrices:	bold capital
vectors:	bold lowercase
scalars, integers:	lowercase

Symbols

<i>Symbol</i>	<i>Description</i>
O	an object or stimulus
m	the number of objects, stimuli
i	index $i = 1, \dots, m$
j	index $j = 1, \dots, m$
k	index $k = 1, \dots, m$
n	the number of object pairs = $\frac{1}{2}m(m-1)$
l	index $l = 1, \dots, n$
N	the number of replications of samples of size $n \times 1$
ℓ	index $\ell = 1, \dots, N$
f	a frequency value associated with an object pair
δ	a dissimilarity value associated with an object pair
$\hat{\delta}$	an estimated dissimilarity value associated with an object pair
$\boldsymbol{\delta}$	an $n \times 1$ vector with dissimilarities between all object pairs
$\hat{\boldsymbol{\delta}}$	an $n \times 1$ vector with estimated dissimilarities between all object pairs
$\mathbf{\Delta}$	an $m \times m$ matrix with dissimilarities
$\tilde{\Delta}_{l\ell}$	a random variable producing realisations $\tilde{\delta}_{l\ell}$
$\tilde{\delta}_{l\ell}$	a realisation of random variable $\tilde{\Delta}_{l\ell}$
$\tilde{\mathbf{\Delta}}$	an $n \times N$ matrix of random variables $\tilde{\Delta}_{l\ell}$
$\bar{\Delta}_l$	mean of a row l of $\mathbf{\Delta}$
ζ	a similarity value associated with an object pair
$\boldsymbol{\zeta}$	an $n \times 1$ vector with similarities between all object pairs
$\boldsymbol{\Sigma}_{\zeta}$	an $m \times m$ matrix with similarities
F	a feature, which is a binary $(0, 1)$ vector of size $m \times 1$
F_C	a cluster feature, which is a binary $(0, 1)$ vector of size $m \times 1$
F_U	a unique feature, which is a binary $(0, 1)$ vector of size $m \times 1$
T	the number of features

T_C	the number of <i>cluster</i> features
T_U	the number of <i>unique</i> features
T_D	the total number of distinctive features = $\frac{1}{2}(2^m) - 1$
t	index for the features: $t = 1, \dots, T$
t_C	index for the <i>cluster</i> features: $t_C = 1, \dots, T_C$
t_U	index for the <i>unique</i> features: $t_U = 1, \dots, T_U$
S_i	the set of features that represents object O_i
\mathbf{E}	an $m \times T$ matrix with columns representing features
\mathbf{e}	a row vector from the matrix \mathbf{E}
e	an element of the matrix \mathbf{E}
\mathbf{E}_T	an \mathbf{E} matrix with special feature structure that yields a tree representation
\mathbf{E}_C	the part of \mathbf{E}_T (size $m \times T_C$) that represents the set of <i>cluster</i> features
\mathbf{E}_U	the part of \mathbf{E}_T (size $m \times T_U$) that represents the set of <i>unique</i> features
\mathbf{X}	an $n \times T$ matrix with featurewise distances obtained with $\mathbf{x}' = \mathbf{e}_{it} - \mathbf{e}_{jt} $
\mathbf{x}'	a row vector from the matrix \mathbf{X}
\mathbf{x}	a column vector from the matrix \mathbf{X}
\mathbf{X}_T	an $n \times T_C + T_U$ matrix with featurewise distances obtained with \mathbf{E}_T
\mathcal{D}	the complete set of featurewise distances
d	a distance between an object pair
\mathbf{d}	an $n \times 1$ vector of distances between all object pairs
$\hat{\mathbf{d}}$	an $n \times 1$ vector of estimated distances between all object pairs
$\hat{\mathbf{d}}_T$	an $n \times 1$ vector of estimated distances between all object pairs for a tree structure
η	feature discriminability parameter
η_{OLS}	true value of ordinary least squares feature discriminability parameter
η_{ICLS}	true value of inequality constrained least squares feature discriminability parameter
η_L	true value of Lasso feature discriminability parameter
η_{PL}	true value of Positive Lasso feature discriminability parameter
$\boldsymbol{\eta}$	an $T \times 1$ vector of feature discriminability parameters
$\boldsymbol{\eta}_{OLS}$	an $T \times 1$ vector of true values η_{OLS}
$\boldsymbol{\eta}_{ICLS}$	an $T \times 1$ vector of true values η_{ICLS}
$\boldsymbol{\eta}_L$	an $T \times 1$ vector of true values η_L
$\boldsymbol{\eta}_{PL}$	an $T \times 1$ vector of true values η_{PL}
$\hat{\eta}, \hat{\eta}_{OLS}$	estimated values of $\eta, \eta_{OLS}, \eta_{ICLS}, \eta_L, \eta_{PL}$
C	the number of constraints necessary to obtain $\hat{\boldsymbol{\eta}}_{ICLS}$
c	index $c = 1, \dots, C$
\mathbf{r}	a $C \times 1$ vector with constraints
\mathbf{A}	a $C \times T$ matrix of constraints of rank c
$\boldsymbol{\lambda}_{KT}$	a $m \times 1$ vector with Kuhn-Tucker multipliers
$\boldsymbol{\epsilon}$	a $n \times 1$ vector with error values ($\boldsymbol{\epsilon} = \boldsymbol{\delta} - \mathbf{X}\boldsymbol{\eta}$)
$\hat{\boldsymbol{\epsilon}}$	a $n \times 1$ vector with estimated error values ($\hat{\boldsymbol{\epsilon}} = \boldsymbol{\delta} - \mathbf{X}\hat{\boldsymbol{\eta}}$)
$\hat{\epsilon}$	an element from the vector $\hat{\boldsymbol{\epsilon}}$
σ^2, σ	true variance and standard deviation of $\boldsymbol{\epsilon}$
$\hat{\sigma}^2, \hat{\sigma}$	estimated variance and standard deviation of $\hat{\boldsymbol{\epsilon}}$
$\sigma_{\eta}^2, \sigma_{\eta}$	true variance and standard error of η
$\hat{\sigma}_{\eta}^2, \hat{\sigma}_{\eta}$	estimated nominal variance and estimated nominal standard error of η
$\hat{\sigma}_{\hat{\eta}}^2, \hat{\sigma}_{\hat{\eta}}$	estimated nominal variance and nominal standard error of $\hat{\eta}$
$\sigma_{OLS}^2, \sigma_{OLS}$	true variance and standard error of $\hat{\eta}_{OLS}$

$\hat{\sigma}_{\text{OLS}}^2, \hat{\sigma}_{\text{OLS}}$	estimated variance and standard error of $\hat{\eta}_{\text{OLS}}$
$\sigma_{\text{ICLS}}^2, \sigma_{\text{ICLS}}$	true variance and standard error of $\hat{\eta}_{\text{ICLS}}$
$\hat{\sigma}_{\text{ICLS}}^2, \hat{\sigma}_{\text{ICLS}}$	estimated variance and standard error of $\hat{\eta}_{\text{ICLS}}$
B	number of bootstrap samples
b	index $b = 1, \dots, B$
\mathbf{b}_b	a bootstrap sample ($n \times 1$ vector)
\mathbf{b}_b^*	a bootstrap sample, multivariate
$\tilde{\mathbf{b}}_b$	a bootstrap sample, with sampled residuals
sd_B	standard deviation of B bootstrap samples
S	number of simulation samples
a	index $a = 1, \dots, S$
\mathbf{s}^*	a simulation sample ($n \times 1$ vector)
κ, p	parameters binomial distribution
GCV	generalized cross-validation statistic
GCV_{FNM}	GCV using inequality constrained least squares estimation

