

Classification Methods for Multi-Class Multivariate Longitudinal Data

José Cortiñas Abrahantes

<mailto:jose.cortinas@uhasselt.be>

205 Inserm Workshop



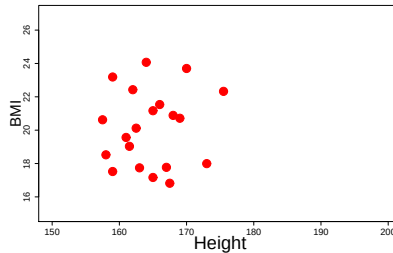
June 2 – 4, 2010



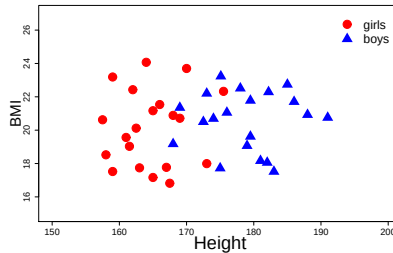
Outline

- Introduction
- Psychotropic drugs – EEG data
- Doubly Hierarchical Supervised Learning Analysis
 - Phase I: Data Modelling
 - Phase II: Supervised Learning Analysis
 - Model Averaging
 - Application: EEG study
- Multivariate Functional Linear Discriminant Analysis
 - Methodology
 - Application: EEG study
- Conclusion

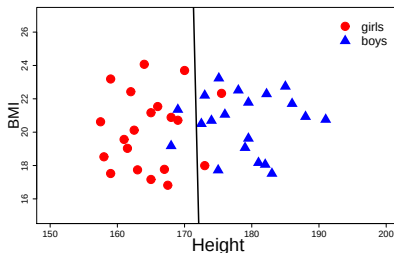
Introduction – Classical Classification Problem



Introduction – Classical Classification Problem

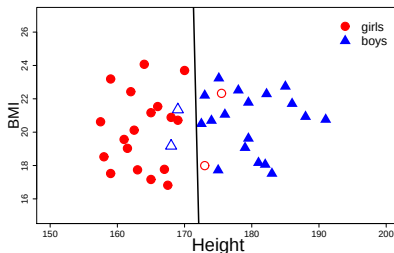


Introduction – Classical Classification Problem



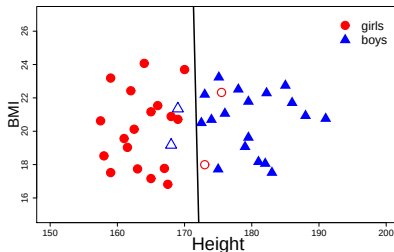
- Linear Discriminant Analysis (LDA):
Find optimal line that separates classes

Introduction – Classical Classification Problem



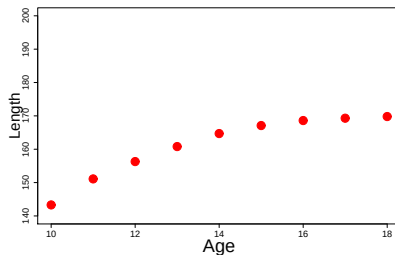
- Linear Discriminant Analysis (LDA):
Find optimal line that separates classes
- Misclassification: 2 boys and 2 girls

Introduction – Classical Classification Problem



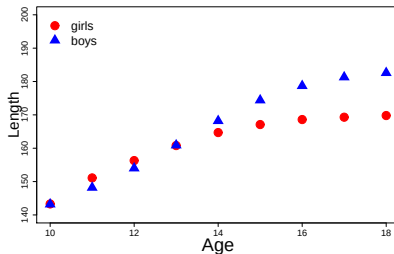
- Linear Discriminant Analysis (LDA):
Find optimal line that separates classes
- Misclassification: 2 boys and 2 girls
- Other supervised learning techniques:
 - Mixture DA, Flexible DA, Penalized DA, Non-parametric DA,...
 - Random Forests, Support Vector Machines, Distance Weighted Discrimination,...

Introduction – Longitudinal Data



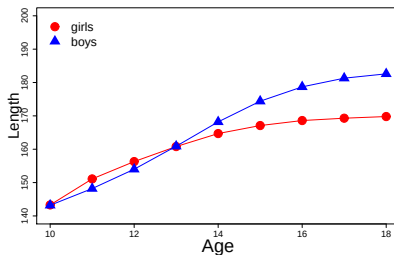
- Growth curve : Evolution of length over time

Introduction – Longitudinal Data



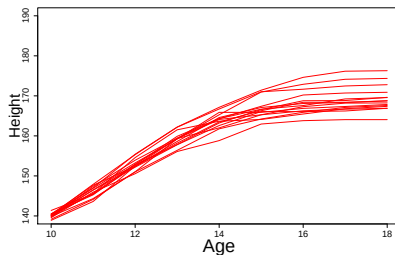
- Growth curve : Evolution of length over time
 - Different evolution for boys and girls

Introduction – Longitudinal Data



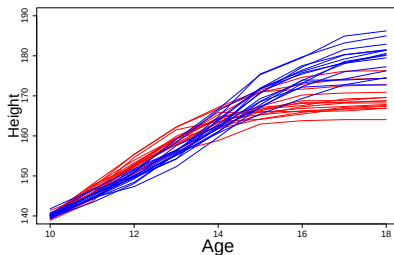
- Growth curve : Evolution of length over time
 - Different evolution for boys and girls
 - Measurements from same boy/girl are correlated

Introduction – Longitudinal Data



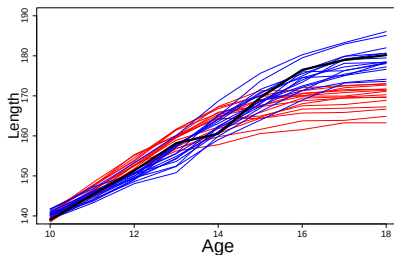
- Growth curve : Evolution of length over time
 - Different evolution for boys and girls
 - Measurements from same boy/girl are correlated

Introduction – Longitudinal Data



- Growth curve : Evolution of length over time
 - Different evolution for boys and girls
 - Measurements from same boy/girl are correlated

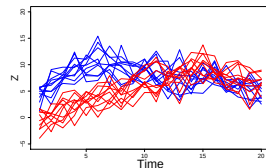
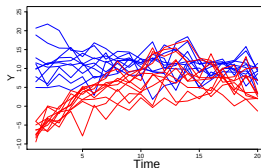
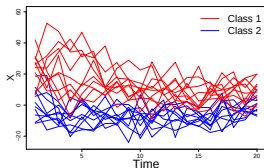
Introduction – Longitudinal Data



- Growth curve : Evolution of length over time
 - Different evolution for boys and girls
 - Measurements from same boy/girl are correlated

Introduction - Research Question

**How can we classify multi-class
multivariate longitudinal data?**



Case Study - Classification of Psychotropic Drugs

Psychotropic Drugs

- What?

Substances that act on the central nervous system and affect mood, thinking and behaviour.

- Classification

Named according to its main indication in psychiatry

- Antidepressants
- Antipsychotics
- Hypnotics
- Stimulants

Case Study - Classification of Psychotropic Drugs

What can we use to classify psychotropic drugs?

- EEG
 - Measure brain activity of a rat
 - Sleep Wake cycle
- 6 sleeping stages:
 - Active Wake (AW)
 - Passive Wake (PW)
 - Light Sleep (SWS1)
 - Deep Sleep (SWS2)
 - Intermediate Stage (IS)
 - REM Sleep (RS)

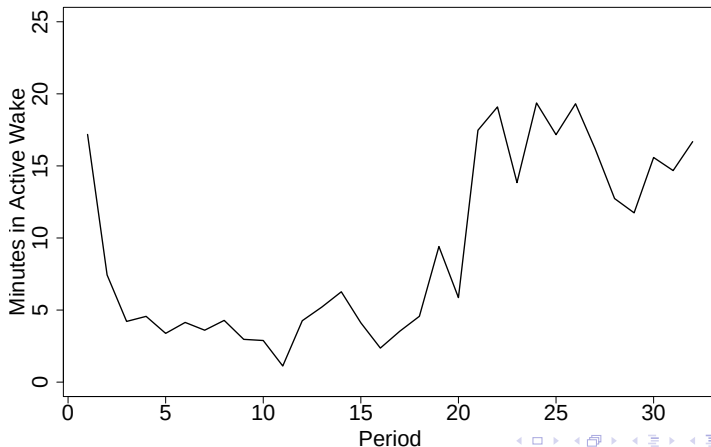
Case Study - Classification of Psychotropic Drugs

Setting of our experiment

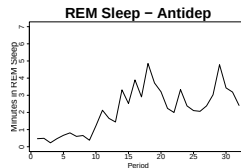
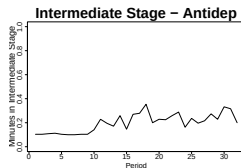
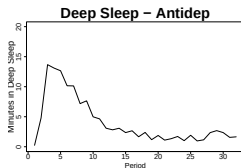
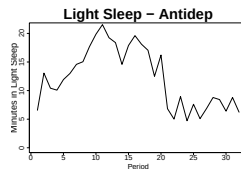
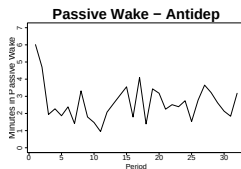
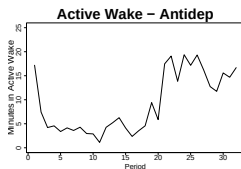
- 26 psychoactive agents
- 4 doses (incl. dose 0) – 8 rats per dose
- 16h → Light period (10h) and Dark period (6h)
- For every period of 30 minutes, number of minutes spent in each sleeping stage is recorded.

Case Study - Classification of Psychotropic Drugs

Active Wake - Antidepressant



Case Study - Classification of Psychotropic Drugs



Case Study - Classification of Psychotropic Drugs

Setting of our experiment

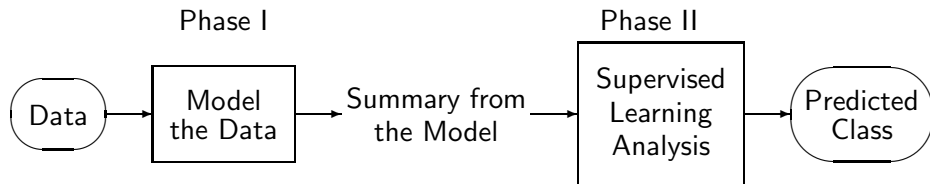
- 26 psychoactive agents
- 4 doses (incl. dose 0) – 8 rats per dose
- 16h → Light period (10h) and Dark period (6h)
- For every period of 30 minutes, number of minutes spent in each sleeping stage is recorded.

Case Study - Classification of Psychotropic Drugs

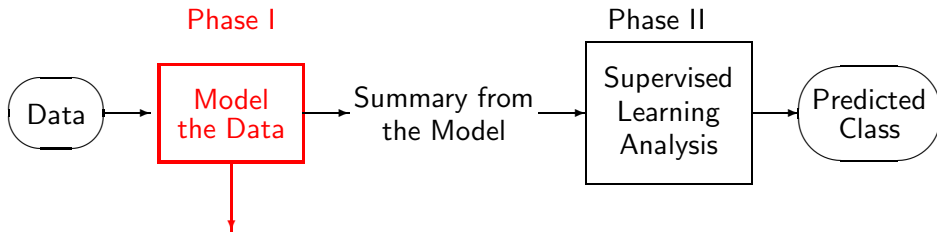
Setting of our experiment

- 26 psychoactive agents
- 4 doses (incl. dose 0) – 8 rats per dose
- 16h → Light period (10h) and Dark period (6h)
- For every period of 30 minutes, number of minutes spent in each sleeping stage is recorded.
- **Training data set:** 58 compound-dose combinations
 - 23 placebo, 13 antidepressants, 7 antipsychotics, 5 hypnotics, 10 stimulants
- **Test data set:** 14 compound-dose combinations
 - 3 placebo, 4 antidepressants, 2 antipsychotics, 2 hypnotics, 3 stimulants

Doubly Hierarchical Supervised Learning Analysis

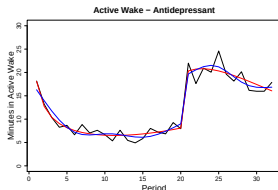


Doubly Hierarchical Supervised Learning Analysis



Fractional Polynomial Mixed Model

Splines Model



Phase I - Fractional Polynomial Mixed Model

Second degree fractional polynomial mixed model (Royston & Altman, 1994)

$$Y_{ijk} = \beta_{0i} + \beta_{1i} \cdot H_1(t_{ijk}) + \beta_{2i} \cdot H_2(t_{ijk}) + \epsilon_{ijk}$$

$$\text{where } H_r(X) = \begin{cases} X^{(p_r)} & p_r \neq p_{r-1} \\ H_{r-1}(X) \cdot \ln(X) & p_r = p_{r-1} \end{cases}$$

and p_1, p_2 selected from $\{-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2\}$
based on likelihood

Phase I - Fractional Polynomial Mixed Model

Second degree fractional polynomial mixed model (Royston & Altman, 1994)

$$Y_{ijk} = (\beta_{0i} + b_{0ij}) + (\beta_{1i} + b_{1ij}) \cdot H_1(t_{ijk}) + (\beta_{2i} + b_{2ij}) \cdot H_2(t_{ijk}) + \epsilon_{ijk}$$

$$\text{where } H_r(X) = \begin{cases} X^{(p_r)} & p_r \neq p_{r-1} \\ H_{r-1}(X) \cdot \ln(X) & p_r = p_{r-1} \end{cases}$$

and p_1, p_2 selected from $\{-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2\}$
based on likelihood

Phase I - Fractional Polynomial Mixed Model

Second degree fractional polynomial mixed model (Royston & Altman, 1994)

$$Y_{ijk} = (\beta_{0i} + b_{0ij}) + (\beta_{1i} + b_{1ij}) \cdot H_1(t_{ijk}) + (\beta_{2i} + b_{2ij}) \cdot H_2(t_{ijk}) + \epsilon_{ijk}$$

$$\text{where } H_r(X) = \begin{cases} X^{(p_r)} & p_r \neq p_{r-1} \\ H_{r-1}(X) \cdot \ln(X) & p_r = p_{r-1} \end{cases}$$

- Fit model per treatment for each sleeping stage
- Separate model for light period, dark period and first three hours

Phase I - Fractional Polynomial Mixed Model

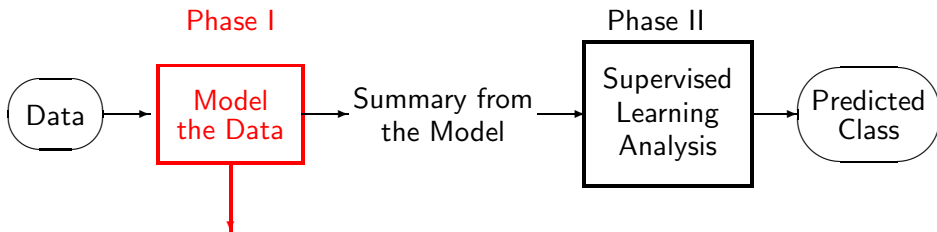
Second degree fractional polynomial mixed model (Royston & Altman, 1994)

$$Y_{ijk} = (\beta_{0i} + b_{0ij}) + (\beta_{1i} + b_{1ij}) \cdot H_1(t_{ijk}) + (\beta_{2i} + b_{2ij}) \cdot H_2(t_{ijk}) + \epsilon_{ijk}$$

$$\text{where } H_r(X) = \begin{cases} X^{(p_r)} & p_r \neq p_{r-1} \\ H_{r-1}(X) \cdot \ln(X) & p_r = p_{r-1} \end{cases}$$

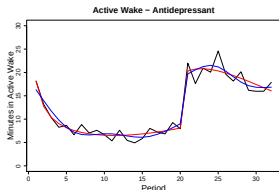
- Fit model per treatment for each sleeping stage
- Separate model for light period, dark period and first three hours
 - light period $\beta_{0i} + b_{0ij}, \beta_{1i} + b_{1ij}, \beta_{2i} + b_{2ij}, p_{1il}, p_{2il}$
 - dark period $\gamma_{0i} + c_{0ij}, \gamma_{1i} + c_{1ij}, \gamma_{2i} + c_{2ij}, p_{1id}, p_{2id}$
 - first 3 hours $\delta_{0i} + d_{0ij}, \delta_{1i} + d_{1ij}, \delta_{2i} + d_{2ij}, p_{1if}, p_{2if}$

Doubly Hierarchical Supervised Learning Analysis

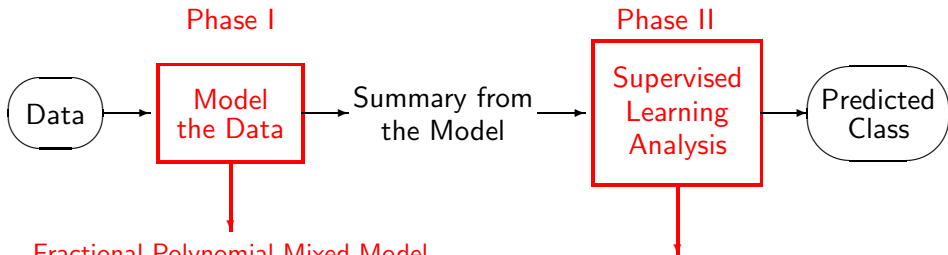


Fractional Polynomial Mixed Model

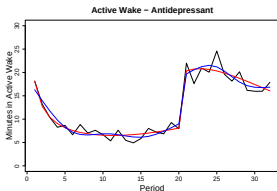
Splines Model



Doubly Hierarchical Supervised Learning Analysis



Fractional Polynomial Mixed Model
Splines Model



Stepwise Linear Discriminant Analysis
JRSS Series C, 56 (2), 223–234.

Stepwise Flexible Discriminant Analysis
Stepwise Mixture Discriminant Analysis
Biopharmaceutical Statistics, 08

Phase II - Linear Discriminant Analysis

- Assumption: Class $c \sim N(\mu_c, \Sigma)$
- Density for class c

$$\begin{aligned} f_c(x) &= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu_c)^T \Sigma^{-1} (\mathbf{x} - \mu_c) \right) \\ &\sim \exp \left[\sum_{j=1}^p \alpha_{jc} x_j \right] \end{aligned}$$

Phase II - Linear Discriminant Analysis

- Assumption: Class $c \sim N(\mu_c, \Sigma)$
- Density for class c

$$\begin{aligned} f_c(x) &= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu_c)^T \Sigma^{-1} (\mathbf{x} - \mu_c) \right) \\ &\sim \exp \left[\sum_{j=1}^p \alpha_{jc} x_j \right] \end{aligned}$$

- Example: 2 classes

$$(\mu_1 - \mu_2)^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left[\frac{c(1|2) p_2}{c(2|1) p_1} \right]$$

→ Linear classification rule!

→ Easy interpretation

Phase II - Stepwise Linear Discriminant Analysis

- Stepwise procedure, where subsequently one drugclass is differentiated from the rest.

Step 1: Stimulants

Step 2: Antipsychotics

Step 3: Antidepressants

Step 4: Hypnotics

Phase II - Stepwise Linear Discriminant Analysis

- Stepwise procedure, where subsequently one drugclass is differentiated from the rest.

$$\begin{aligned}\text{Step 1: Stimulants} &\Rightarrow P_1 \\ \text{Step 2: Antipsychotics} &\Rightarrow P_2(1 - P_1) \\ \text{Step 3: Antidepressants} &\Rightarrow P_3(1 - P_2)(1 - P_1) \\ \text{Step 4: Hypnotics} &\Rightarrow P_4(1 - P_3)(1 - P_2)(1 - P_1)\end{aligned}$$

- The posterior probabilities are calculated hierarchically

Phase II - Stepwise Linear Discriminant Analysis

- Stepwise procedure, where subsequently one drugclass is differentiated from the rest.

Step 1: Stimulants $\Rightarrow P_1$

Step 2: Antipsychotics $\Rightarrow P_2(1 - P_1)$

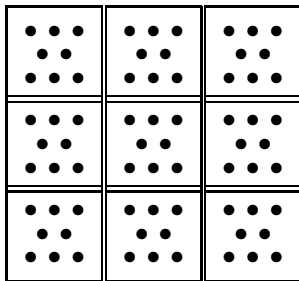
Step 3: Antidepressants $\Rightarrow P_3(1 - P_2)(1 - P_1)$

Step 4: Hypnotics $\Rightarrow P_4(1 - P_3)(1 - P_2)(1 - P_1)$

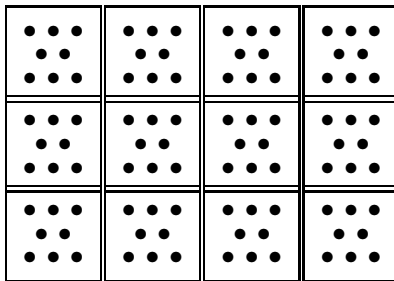
- The posterior probabilities are calculated hierarchically
- Model selection: 10-fold cross validation
 - Selection Procedure I: rat level
 - Selection Procedure II: compound-dose level

Model selection: 10-fold cross validation

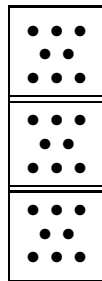
Class 1



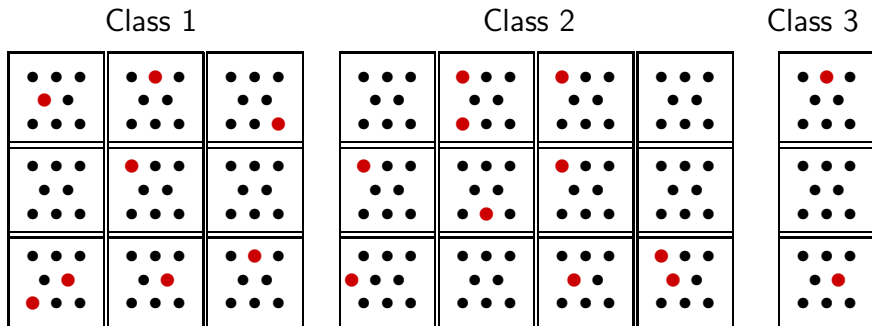
Class 2



Class 3

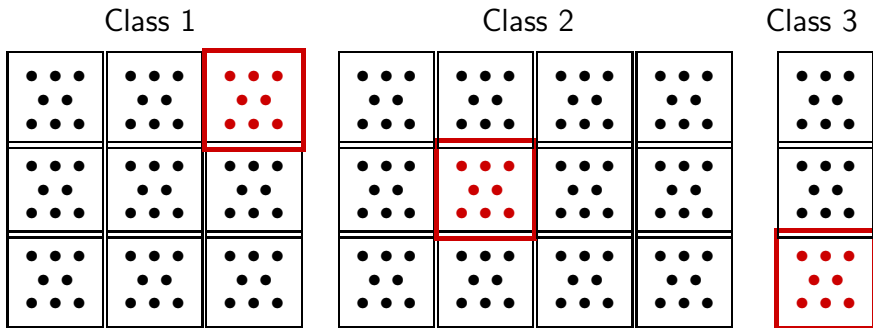


Model selection: 10-fold cross validation



Rat Level

Model selection: 10-fold cross validation



Compound-dose Level

Phase II - Variable Selection Criterium

- Define two error measures:

$$\begin{aligned}\text{Error1}_s &= \text{ERR}_{C_s C_{-s}} + (1 - \text{PP}_{C_s C_s}) \\ \text{Error2}_s &= \text{ERR}_{C_{-s} C_s} + \sum_{k \neq C_s} \text{PP}_{k C_s}\end{aligned}$$

where

- g = initial number of classes
- ERR_{kl} = Misclassification percentage from class k in class l
- PP_{kl} = Posterior probability for in class k to be classified in class l
- C_s = Class discriminated in step s
- C_{-s} = Other classes retained in the dataset in step s

Phase II - Variable Selection Criterium

- Define two error measures:

$$\begin{aligned}\text{Error1}_s &= \text{ERR}_{C_s C_{-s}} + (1 - \text{PP}_{C_s C_s}) \\ \text{Error2}_s &= \text{ERR}_{C_{-s} C_s} + \sum_{k \neq C_s} \text{PP}_{k C_s}\end{aligned}$$

where

- g = initial number of classes
- ERR_{kl} = Misclassification percentage from class k in class l
- PP_{kl} = Posterior probability for in class k to be classified in class l
- C_s = Class discriminated in step s
- C_{-s} = Other classes retained in the dataset in step s

- Lack of classification measure (LC) in step s

$$\text{LC}_s = (s + 1) \cdot \text{Error1}_s + 2 \cdot (g - s) \cdot \text{Error2}_s$$

Results - Selection Procedure I

Training Dataset – Error count = 0.000

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.97	0.02	0.01	0.00	0.00
Antipsychotic	0.01	0.94	0.03	0.00	0.02
Antidepressant	0.00	0.03	0.95	0.01	0.01
Hypnotic	0.00	0.00	0.01	0.99	0.00
Stimulant	0.01	0.01	0.02	0.00	0.96

Validation Dataset – Error count = 0.600

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.96	0.00	0.04	0.00	0.00
Antipsychotic	0.38	0.33	0.03	0.26	0.00
Antidepressant	0.11	0.45	0.16	0.16	0.12
Hypnotic	0.48	0.03	0.31	0.18	0.00
Stimulant	0.01	0.23	0.07	0.04	0.65

Results - Selection Procedure II

Training Dataset – Error count = 0.043

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.94	0.01	0.01	0.04	0.00
Antipsychotic	0.00	0.72	0.10	0.11	0.07
Antidepressant	0.00	0.18	0.75	0.06	0.01
Hypnotic	0.00	0.04	0.18	0.78	0.00
Stimulant	0.04	0.04	0.09	0.04	0.79

Results - Selection Procedure II

Training Dataset – Error count = 0.043

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.94	0.01	0.01	0.04	0.00
Antipsychotic	0.00	0.72	0.10	0.11	0.07
Antidepressant	0.00	0.18	0.75	0.06	0.01
Hypnotic	0.00	0.04	0.18	0.78	0.00
Stimulant	0.04	0.04	0.09	0.04	0.79

Validation Dataset – Error count = 0.600

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.33	0.00	0.05	0.62	0.00
Antipsychotic	0.24	0.33	0.00	0.43	0.00
Antidepressant	0.02	0.38	0.49	0.02	0.09
Hypnotic	0.93	0.02	0.02	0.03	0.00
Stimulant	0.00	0.03	0.28	0.00	0.69

Discussion

- Other techniques can be used in Phase I and Phase II
- Nice classification in training dataset
- Classification results obtained in training dataset are slightly better for Selection procedure I
 - Few compound-dose combinations in some classes
 - Different aspect of variability

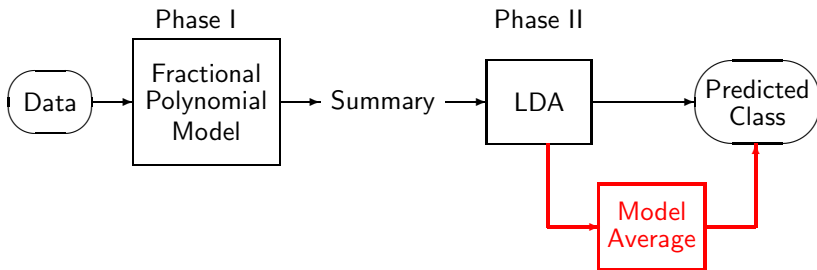
Discussion

- Other techniques can be used in Phase I and Phase II
- Nice classification in training dataset
- Classification results obtained in training dataset are slightly better for Selection procedure I
 - Few compound-dose combinations in some classes
 - Different aspect of variability
- Poor classification in test dataset

Discussion

- Other techniques can be used in Phase I and Phase II
- Nice classification in training dataset
- Classification results obtained in training dataset are slightly better for Selection procedure I
 - Few compound-dose combinations in some classes
 - Different aspect of variability
- Poor classification in test dataset
 - Model selection bias?

Doubly Hierarchical Supervised Learning Analysis



General idea (Burnham and Anderson, 2002)

- Consider the N best models
- Calculate weighted average of the coefficients
- Use these weighted coefficients to obtain a new discriminant rule

Details

$$\begin{aligned} \text{Calculate weights } w_s^{(m)} &= \frac{\exp(-\Delta_s^{(m)}/2)}{\sum_{r=1}^R \exp(-\Delta_s^{(r)}/2)} \\ \Delta_s^{(m)} &= \text{LC}_s^{(m)} - \min_r(\text{LC}_s^{(r)}) \end{aligned}$$

$$\text{The coefficients are then } \hat{\alpha}_{j,c} = \frac{\sum_{m=1}^R w^{(m)} l_j(m) \hat{\alpha}_{j,c}^{(m)}}{w_+(j)}$$

$$\begin{aligned} \text{where } w_+(j) &= \sum_{m=1}^R w^{(m)} l_j(m) \\ l_j(m) &= \begin{cases} 1 & \text{if predictor } x_j \text{ is in model } m, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Model Averaging: Selection Procedure II

Without model averaging – error count = 0.600

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.33	0.01	0.05	0.61	0.00
Antipsychotic	0.24	0.34	0.00	0.42	0.00
Antidepressant	0.02	0.38	0.49	0.02	0.09
Hypnotic	0.93	0.02	0.02	0.03	0.00
Stimulant	0.01	0.02	0.28	0.00	0.69

Model Averaging: Selection Procedure II

Without model averaging – error count = 0.600

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.33	0.01	0.05	0.61	0.00
Antipsychotic	0.24	0.34	0.00	0.42	0.00
Antidepressant	0.02	0.38	0.49	0.02	0.09
Hypnotic	0.93	0.02	0.02	0.03	0.00
Stimulant	0.01	0.02	0.28	0.00	0.69

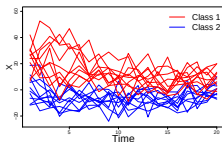
Using 25 models – error count = 0.400

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.89	0.00	0.11	0.00	0.00
Antipsychotic	0.32	0.62	0.02	0.04	0.00
Antidepressant	0.05	0.51	0.21	0.12	0.11
Hypnotic	0.70	0.03	0.20	0.07	0.00
Stimulant	0.01	0.08	0.20	0.00	0.71

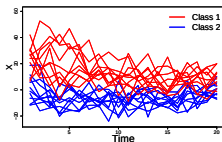
Discussion: Model Averaging

- Performance evaluated using training-test setting
- Use of more than one classification rule (model) improved results
- Error count in the test data, reduced more than 50 %
- Posterior probabilities increases
- No need to use more than 100 models

Functional Data Classification

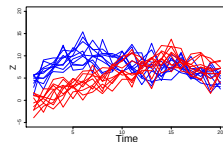
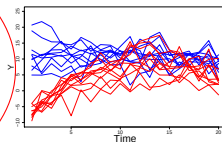
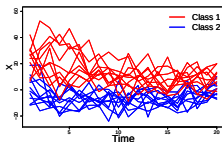


Functional Data Classification



Functional Linear
Discriminant Analysis
(James and Hastie, 2001)

Functional Data Classification



Functional Linear
Discriminant Analysis → Extension to Multivariate Case
(James and Hastie, 2001)

Univariate Functional Linear Discriminant Analysis

- James and Hastie (2001)
- Fit mixed model with splines as random and fixed effect

$$\mathbf{Y}_{ic} = S\beta_c + S\mathbf{b}_{ic} + \varepsilon_{ic}$$

with S natural cubic spline basis with K knots

$$b_{ic} \sim N(\mathbf{0}, D_c)$$

$$\varepsilon_{ic} \sim N(\mathbf{0}, \Sigma_{ic})$$

- Marginal model: $\mathbf{Y}_{ic} \sim N(S\beta_c, \Sigma_{ic} + SD_cS^T)$
- Classification rule for new curve $\mathbf{y}$

$$d(\mathbf{y}, c) = \sqrt{(\mathbf{y} - S\beta_c)(\Sigma_c + SD_cS^T)^{-1}(\mathbf{y} - S\beta_c)^T}$$

$\mathbf{y}$ is classified to class c for which distance $d(\mathbf{y}, c)$ is minimal.

Pairwise Pseudo-likelihood Modeling

- Fieuws and Verbeke (2006)
- Assume m outcomes $Y_1, \dots, Y_m$
- For each pair (Y_r, Y_s) , we fit the model:

$$\begin{pmatrix} Y_{ri} \\ Y_{si} \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \beta_r \\ \beta_s \end{pmatrix} + \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} D_r & D_{rs} \\ D_{rs} & D_s \end{pmatrix} \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_r & \Sigma_{rs} \\ \Sigma_{rs} & \Sigma_s \end{pmatrix} \right)$$

$$(r = 1, \dots, m-1; s = r+1, \dots, m)$$

Pairwise Pseudo-likelihood Modeling

- Fieuws and Verbeke (2006)
- Assume m outcomes $Y_1, \dots, Y_m$
- For each pair (Y_r, Y_s) , we fit the model:

$$\begin{pmatrix} Y_{ri} \\ Y_{si} \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \beta_r \\ \beta_s \end{pmatrix} + \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} D_r & D_{rs} \\ D_{rs} & D_s \end{pmatrix} \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_r & \Sigma_{rs} \\ \Sigma_{rs} & \Sigma_s \end{pmatrix} \right)$$

$$(r = 1, \dots, m-1; s = r+1, \dots, m)$$

Pairwise Pseudo-likelihood Modeling

- Fieuws and Verbeke (2006)
- Assume m outcomes $Y_1, \dots, Y_m$
- For each pair (Y_r, Y_s) , we fit the model:

$$\begin{pmatrix} Y_{ri} \\ Y_{si} \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \beta_r \\ \beta_s \end{pmatrix} + \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{ri} \\ \mathbf{b}_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} D_r & D_{rs} \\ D_{rs} & D_s \end{pmatrix} \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{ri} \\ \varepsilon_{si} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_r & \Sigma_{rs} \\ \Sigma_{rs} & \Sigma_s \end{pmatrix} \right)$$

$$(r = 1, \dots, m-1; s = r+1, \dots, m)$$

Pairwise Pseudo-likelihood Modeling

$$\begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix} = \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \sum \frac{\beta_1}{m-1} \\ \vdots \\ \sum \frac{\beta_m}{m-1} \end{pmatrix} + \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum \frac{D_1}{m-1} & D_{12} & \cdots & D_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{1m} & D_{2m} & \cdots & \sum \frac{D_m}{m-1} \end{pmatrix} \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum \frac{\Sigma_1}{m-1} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1m} & \Sigma_{2m} & \cdots & \sum \frac{\Sigma_m}{m-1} \end{pmatrix} \right)$$

Pairwise Pseudo-likelihood Modeling

$$\begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix} = \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_m^* \end{pmatrix} + \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum \frac{D_1}{m-1} & D_{12} & \cdots & D_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{1m} & D_{2m} & \cdots & \sum \frac{D_m}{m-1} \end{pmatrix} \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum \frac{\Sigma_1}{m-1} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1m} & \Sigma_{2m} & \cdots & \sum \frac{\Sigma_m}{m-1} \end{pmatrix} \right)$$

Pairwise Pseudo-likelihood Modeling

$$\begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix} = \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_m^* \end{pmatrix} + \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \mathbf{D}^* \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum \frac{\Sigma_1}{m-1} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1m} & \Sigma_{2m} & \cdots & \sum \frac{\Sigma_m}{m-1} \end{pmatrix} \right)$$

Pairwise Pseudo-likelihood Modeling

$$\begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix} = \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \vdots \\ \beta_m^* \end{pmatrix} + \begin{pmatrix} S & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{pmatrix} \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{1i} \\ \vdots \\ \mathbf{b}_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \mathbf{D}^* \right)$$

$$\text{and } \begin{pmatrix} \varepsilon_{1i} \\ \vdots \\ \varepsilon_{mi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \mathbf{\Sigma}^* \right)$$

Multivariate Functional Linear Discriminant Analysis

- Marginal model:

$$\mathbf{Y}_{ic} \sim N(S\beta_c^*, \Sigma_c^* + SD_c^*S^T)$$

Multivariate Functional Linear Discriminant Analysis

- Marginal model:

$$\mathbf{Y}_{ic} \sim N(S\beta_c^*, \Sigma_c^* + SD_c^* S^T)$$

- Classification rule for new curve $\mathbf{y}$

$$d(\mathbf{y}, c) = \sqrt{(\mathbf{y} - S\beta_c^*)(\Sigma_c^* + SD_c^* S^T)^{-1}(\mathbf{y} - S\beta_c^*)^T}$$

$\mathbf{y}$ is classified to class c for which distance $d(\mathbf{y}, c)$ is minimal.

Multivariate Functional Linear Discriminant Analysis

- Marginal model:

$$\mathbf{Y}_{ic} \sim N(S\beta_c^*, \Sigma_c^* + SD_c^* S^T)$$

- Classification rule for new curve $\mathbf{y}$

$$d(\mathbf{y}, c) = \sqrt{(\mathbf{y} - S\beta_c^*) \Gamma_c^{*-1} (\mathbf{y} - S\beta_c^*)^T}$$

$\mathbf{y}$ is classified to class c for which distance $d(\mathbf{y}, c)$ is minimal.

- Γ_c^* might be **NOT** positive definite

Strategy I

Rousseeuw and Molenberghs (1993)

Modification of Covariance Matrix

- Eigenvalue decomposition of Γ_c^*

$$\Gamma_c^* = P_c \Lambda_c P_c^T$$

- Replace negative values in Λ_c by a small positive value $\rightarrow \Lambda_{c,\text{modif}}$
- Modified covariance matrix $\Gamma_{c,\text{modif}}^*$

$$\Gamma_{c,\text{modif}}^* = P_c \Lambda_{c,\text{modif}} P_c^T$$

- Calculate modified distance

$$d_{\text{modif}}(\mathbf{y}, c) = \sqrt{(\mathbf{y} - S\beta_c^*) \left[\Gamma_{c,\text{modif}}^* \right]^{-1} (\mathbf{y} - S\beta_c^*)^T}.$$

Strategy II

Pairwise Distance Strategy

- Calculate distance for each pair (Y_r, Y_s)

$$d_{rs}(\mathbf{y}, c) = \left(\begin{pmatrix} y_r \\ y_s \end{pmatrix} - \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \beta_{rc} \\ \beta_{sc} \end{pmatrix} \right) \Gamma_{rs,c}^{-1} \left(\begin{pmatrix} y_r \\ y_s \end{pmatrix} - \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \beta_{rc} \\ \beta_{sc} \end{pmatrix} \right)^T$$

with

$$\Gamma_{rs,c} = \begin{pmatrix} \Sigma_{rc} & \Sigma_{rsc} \\ \Sigma_{rsc} & \Sigma_{sc} \end{pmatrix} + \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} D_{rc} & D_{rsc} \\ D_{rsc} & D_{sc} \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}^T$$

- Average distance over all pairs

$$d_{\text{pair}}(\mathbf{y}, c) = \frac{1}{m(m-1)/2} \sum_{r=1}^{m-1} \sum_{s=r+1}^m d_{rs}(\mathbf{y}, c)$$

Simulation Study

- Based on EEG case study

$$\begin{pmatrix} Y_{1ic} \\ Y_{2ic} \\ Y_{3ic} \end{pmatrix} = \begin{pmatrix} S_5 & 0 & 0 \\ 0 & S_5 & 0 \\ 0 & 0 & S_5 \end{pmatrix} \begin{pmatrix} \beta_{1c}^* \\ \beta_{2c}^* \\ \beta_{3c}^* \end{pmatrix} + \begin{pmatrix} S_5 & 0 & 0 \\ 0 & S_5 & 0 \\ 0 & 0 & S_5 \end{pmatrix} \begin{pmatrix} \mathbf{b}_{1ic}^* \\ \mathbf{b}_{2ic}^* \\ \mathbf{b}_{3ic}^* \end{pmatrix} + \begin{pmatrix} \epsilon_{1ic}^* \\ \epsilon_{2ic}^* \\ \epsilon_{3ic}^* \end{pmatrix}$$

$$\text{with } \begin{pmatrix} \mathbf{b}_{1ic}^* \\ \mathbf{b}_{2ic}^* \\ \mathbf{b}_{3ic}^* \end{pmatrix} \sim N(\mathbf{0}, D_c^*) \text{ and } \begin{pmatrix} \epsilon_{1ic}^* \\ \epsilon_{2ic}^* \\ \epsilon_{3ic}^* \end{pmatrix} \sim N(\mathbf{0}, \Sigma_c^*)$$

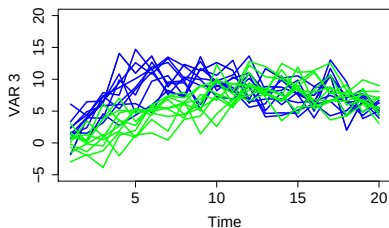
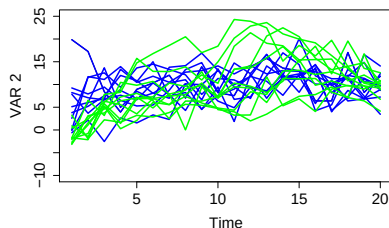
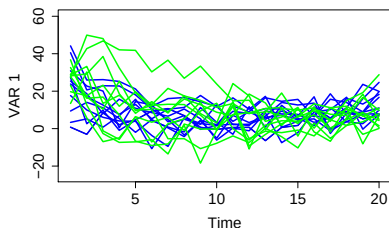
- Generate data from normal distribution:

$$N \left(\begin{pmatrix} S_5 & 0 & 0 \\ 0 & S_5 & 0 \\ 0 & 0 & S_5 \end{pmatrix} \begin{pmatrix} \beta_{1c}^* \\ \beta_{2c}^* \\ \beta_{3c}^* \end{pmatrix}, \Sigma_c^* + \begin{pmatrix} S_5 & 0 & 0 \\ 0 & S_5 & 0 \\ 0 & 0 & S_5 \end{pmatrix} D_c^* \begin{pmatrix} S_5 & 0 & 0 \\ 0 & S_5 & 0 \\ 0 & 0 & S_5 \end{pmatrix} \right)$$

Simulation Setting

- **Setting 1:** Two classes based on Antipsychotics and Stimulants
- **Setting 2:** Three classes based on Antipsychotics, Stimulants and Antidepressants
- **Setting 3:** Four classes based on Antipsychotics, Stimulants, Antidepressants and Hypnotics
- **Training Dataset** → 20 – 40 – 60 – 80 – 100 subjects
- **Test Dataset** → 10 subjects
- Pseudo likelihood model with 2 or 3 knots
- Modified distance ↔ pairwise distance

Illustrative Example



— Class 1
— Class 2

Simulation Study – Classification result

	Number of Subjects	Modified Distance	Pairwise Classification
Setting 1	20	0.035 (0.061)	0.005 (0.019)
	40	0.008 (0.037)	0.002 (0.009)
	60	0.001 (0.007)	0.002 (0.009)
	80	0.001 (0.007)	0.001 (0.005)
	100	0.001 (0.008)	0.002 (0.010)

Simulation Study – Classification result

	Number of Subjects	Modified Distance	Pairwise Classification
Setting 1	20	0.035 (0.061)	0.005 (0.019)
	40	0.008 (0.037)	0.002 (0.009)
	60	0.001 (0.007)	0.002 (0.009)
	80	0.001 (0.007)	0.001 (0.005)
	100	0.001 (0.008)	0.002 (0.010)
Setting 2	20	0.213 (0.074)	0.128 (0.064)
	40	0.151 (0.079)	0.094 (0.058)
	60	0.124 (0.073)	0.076 (0.052)
	80	0.119 (0.069)	0.087 (0.051)
	100	0.092 (0.059)	0.069 (0.048)

Simulation Study – Classification result

	Number of Subjects	Modified Distance	Pairwise Classification
Setting 1	20	0.035 (0.061)	0.005 (0.019)
	40	0.008 (0.037)	0.002 (0.009)
	60	0.001 (0.007)	0.002 (0.009)
	80	0.001 (0.007)	0.001 (0.005)
	100	0.001 (0.008)	0.002 (0.010)
Setting 2	20	0.213 (0.074)	0.128 (0.064)
	40	0.151 (0.079)	0.094 (0.058)
	60	0.124 (0.073)	0.076 (0.052)
	80	0.119 (0.069)	0.087 (0.051)
	100	0.092 (0.059)	0.069 (0.048)
Setting 3	20	0.290 (0.083)	0.251 (0.077)
	40	0.221 (0.084)	0.229 (0.064)
	60	0.194 (0.074)	0.222 (0.060)
	80	0.201 (0.066)	0.215 (0.063)
	100	0.188 (0.069)	0.219 (0.057)

Case Study: Classification Results

Modified Distance – error count = 0.383

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.88	0.08	0.00	0.04	0.00
Antipsychotic	0.00	0.88	0.06	0.00	0.06
Antidepressant	0.19	0.56	0.19	0.03	0.03
Hypnotic	0.67	0.08	0.00	0.25	0.00
Stimulants	0.04	0.33	0.17	0.00	0.46

Case Study: Classification Results

Modified Distance – error count = 0.383

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.88	0.08	0.00	0.04	0.00
Antipsychotic	0.00	0.88	0.06	0.00	0.06
Antidepressant	0.19	0.56	0.19	0.03	0.03
Hypnotic	0.67	0.08	0.00	0.25	0.00
Stimulants	0.04	0.33	0.17	0.00	0.46

Pairwise Distance – error count = 0.367

Drugclass	Placebo	Antipsy	Antidep	Hypnotic	Stimulant
Placebo	0.67	0.08	0.04	0.21	0.00
Antipsychotic	0.00	0.56	0.31	0.00	0.13
Antidepressant	0.22	0.15	0.59	0.03	0.00
Hypnotic	0.42	0.00	0.25	0.33	0.00
Stimulant	0.00	0.13	0.29	0.00	0.58

Discussion: MFLDA

- Extension of Functional Linear Discriminant Analysis to multivariate functional data.
- Similar classification performance of modified distance and pairwise distance.
- Simulation study:
 - Number of classes $\uparrow \rightarrow$ Error rate $\uparrow$
 - Number of subjects $\uparrow \rightarrow$ Error rate $\downarrow$
 - Number of knots $\uparrow \rightarrow$ Error rate $\downarrow$
- EEG dataset:
 - Nice classification results in training dataset,
 - Difficult to classify antidepressants and hypnotics in test dataset

Concluding Remarks

- Classification of multivariate longitudinal data is a non-trivial task
- Need for appropriately tailored methods

Concluding Remarks

- Classification of multivariate longitudinal data is a non-trivial task
- Need for appropriately tailored methods
- **Doubly Hierarchical Supervised Learning Analysis**
 - Phase I: Model the data
 - Phase II: Use summary from the model in a supervised learning analysis
 - Nice classification results in training dataset, but problems to classify test dataset
→ **Model selection bias!**

Concluding Remarks

- Classification of multivariate longitudinal data is a non-trivial task
- Need for appropriately tailored methods
- **Doubly Hierarchical Supervised Learning Analysis**
 - Phase I: Model the data
 - Phase II: Use summary from the model in a supervised learning analysis
 - Nice classification results in training dataset, but problems to classify test dataset
 - **Model selection bias!**
- Model Averaging for Linear Discriminant Analysis
 - Improved classification results in train and test dataset

Concluding Remarks

- **Multivariate Functional Linear Discriminant Analysis**
 - Extension of Functional Linear Discriminant Analysis (James and Hastie, 2001)
 - Pseudo-likelihood modeling for the joint modeling of the curves
 - Non-positive definite variance-covariance matrix
 - Modified distance (Rousseeuw and Molenberghs, 1993)
 - Pairwise distance
 - Nice performance of both methods in training and test dataset, and in simulation studies.

References

- James, G.M., Hastie T.J. (2001). Functional Linear Discriminant Analysis for Irregularly Sampled Curves. *Journal of the Royal Statistical Society. Series B* **63**(3), 533–550.
- Fieuws, S., Verbeke, G. (2006). Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles. *Biometrics* **62**, 424–431.
- Rousseeuw, P.J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics: Theory and Methods*, **22**(4), 965–984.
- Wouters, K., Ahnaou, A., Cortiñas, J., Molenberghs, G., Geys, H., Bijmens, L., Drinkenburg, W.H.I.M. (2007). Psychotropic drug classification based on sleep-wake behaviour of rats. *Journal of the Royal Statistical Society. Series C* **56**(2), 223–234.
- Wouters, K., Cortiñas Abrahantes, J., Molenberghs, G., Ahnaou, A., Bijmens, L., Drinkenburg, W.H.I.M. (2008) A Comparison of Doubly Hierarchical Discriminant Analyses for Multiple Class Longitudinal Data from EEG Experiments. *Journal of Biopharmaceutical Statistics* **18**, 1120–1135.
- Wouters, K., Cortiñas Abrahantes, J., Molenberghs, G., Ahnaou, A., Bijmens, L., Drinkenburg, W.H.I.M. (2009) Correction for Model Selection Bias Using a Modified Model Averaging Approach for Supervised Learning Methods Applied to EEG Experiments. *Conditionally accepted in Journal of Biopharmaceutical Statistics*
- Wouters, K., Cortiñas, J., Molenberghs, G., Geys, H., Ahnaou, A., Drinkenburg, W.H.I.M. and Bijmens, L. (2009). Multivariate Functional Linear Discriminant Analysis Based on Pairwise Pseudo-Likelihood Modeling Combined with Splines. *Manuscript submitted for publication*.

Thank you