PROM International Workshop 2022

A gentle introduction to **Mixture Models** - Part 2

Prof. Salvatore Ingrassia
Department of Economics and Business
University of Catania (Italy)

salvatore.ingrassia@unict.it
http://www.dei.unict.it/docenti/salvatore.ingrassia

Cracow University of Economics
October 6, 2022

2. Mixtures of Regression Models

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts".

Sherlock Holmes, *A Scandal in Bohemia*

Outline

2.1 Mixtures of Regressions

Agenda

- Models with covariates
- Mixture of Regressions (with fixed covariates)
- Mixture Regressions with Concomitant Variables
- Mixture of generalized linear models
- Identifiability of a mixture of linear models

Introduction

Consider a pair $(Y, X_1, \ldots, X_d)$ of a response variable Y and covariates $(X_1, \ldots, X_d)'$ defined on some population Ω with values in $\mathbb{R} \times \mathbb{R}^d$.

Assume we are provided with a sample of N i.i.d. realizations of $(Y, X_1, \ldots, X_d)$ and the dependence of Y_n on $\mathbf{x}_n$ is modeled by a multiple regression model

$$\begin{aligned} Y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_d x_{nd} + \varepsilon_n \\ &= \boldsymbol{\beta}' \mathbf{x}_n + \varepsilon_n \end{aligned}$$

where

- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)' \in \mathbb{R}^{d+1}$ are unknown parameters,
- $\mathbf{x}_n = (1, x_{n1}, \ldots, x_{nd})' \in \mathbb{R}^{d+1}$ denotes the augmented covariate vector.
- $\varepsilon_1, \ldots, \varepsilon_N \sim N(0, \sigma_\varepsilon^2)$.

The problem

In many circumstances, the assumption that the regression coefficients are fixed over all possible realizations of $Y_1, \ldots, Y_N$ is inadequate, and models where the regression coefficients change are of practical interest.

Introduction

Consider a pair $(Y, X_1, \ldots, X_d)$ of a response variable Y and covariates $(X_1, \ldots, X_d)'$ defined on some population Ω with values in $\mathbb{R} \times \mathbb{R}^d$.

Assume we are provided with a sample of N i.i.d. realizations of $(Y, X_1, \ldots, X_d)$ and the dependence of Y_n on $\mathbf{x}_n$ is modeled by a multiple regression model

$$\begin{aligned} Y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_d x_{nd} + \varepsilon_n \\ &= \boldsymbol{\beta}' \mathbf{x}_n + \varepsilon_n \end{aligned}$$

where

- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)' \in \mathbb{R}^{d+1}$ are unknown parameters,
- $\mathbf{x}_n = (1, x_{n1}, \ldots, x_{nd})' \in \mathbb{R}^{d+1}$ denotes the augmented covariate vector.
- $\varepsilon_1, \ldots, \varepsilon_N \sim N(0, \sigma_\varepsilon^2)$.

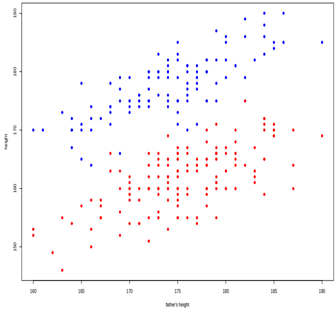
The problem

In many circumstances, the assumption that the regression coefficients are fixed over all possible realizations of $Y_1, \ldots, Y_N$ is inadequate, and models where the regression coefficients change are of practical interest.

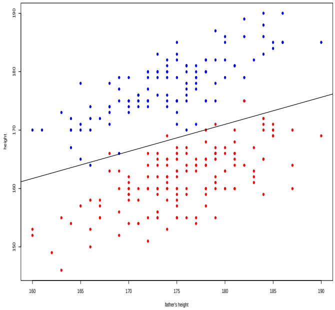
Example A: Student Data

Data come from a survey on $N = 270$ university students.

Consider the relationship between student height and student's father height. Two groups: males and females (blue=males, red=females).



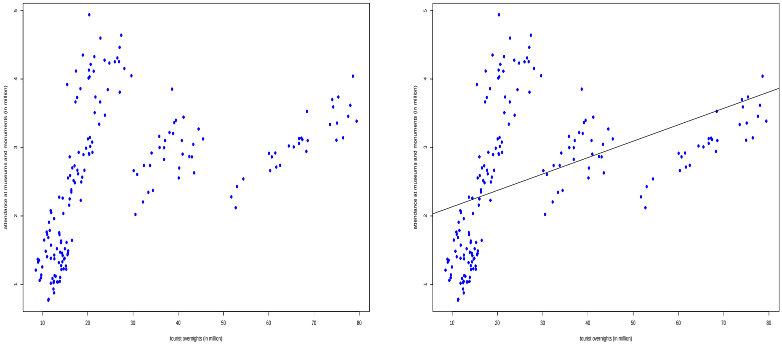
scatter plot



single linear regression model

Example B: Tourism Data

Data concern $N = 180$ monthly observations about the *attendance at museums and monuments* (Y, data in millions) on the *tourist overnights* (X, data in millions) in Italy over the 15-year period spanning from January 1996 to December 2010.



plot of data (*left*); *b)* single linear regression model (*right*).

Mixture models with covariates

Consider a pair $(Y, \mathbf{X})$ of a response variable Y and covariates $\mathbf{X}$ defined on some heterogeneous population Ω partitioned into G disjoint homogeneous subpopulations, i.e. $\Omega = \Omega_1 \cup \cdots \cup \Omega_G$.

We focus on modeling the dependence between Y and $\mathbf{X}$ based on data coming from a heterogeneous population.

In this framework, mixture models provide a flexible approach for a wide variety of random phenomena characterized by unobserved heterogeneity.

Two first approaches

- Mixture of regression models (with fixed covariaters) (MR),
- Mixture of regression models with concomitant variables (MRC)

Mixture models with covariates

Consider a pair $(Y, \mathbf{X})$ of a response variable Y and covariates $\mathbf{X}$ defined on some heterogeneous population Ω partitioned into G disjoint homogeneous subpopulations, i.e. $\Omega = \Omega_1 \cup \cdots \cup \Omega_G$.

We focus on modeling the dependence between Y and $\mathbf{X}$ based on data coming from a heterogeneous population.

In this framework, mixture models provide a flexible approach for a wide variety of random phenomena characterized by unobserved heterogeneity.

Two first approaches

- Mixture of regression models (with fixed covariaters) (MR),
- Mixture of regression models with concomitant variables (MRC)

Mixture of Regression Models (MR) 1/3

Dependence between Y and $\mathbf{X}$ for data coming from a heterogenuous population can be modeled by a *finite mixture of regressions (FMR)*.

Mixtures of Regressions (MR)

Consider a pair $(Y, \mathbf{X})$ of a response variable Y and covariates $\mathbf{X}$ defined on some population Ω. We say that the distribution of $Y|x$ is a mixture of regressions if has pdf given by

$$p(y|\mathbf{x};\boldsymbol{\psi}) = \sum_{g=1}^{G} f(y|\mathbf{x};\boldsymbol{\theta}_g)\pi_g \ .$$

where:

- $f(y|\mathbf{x},\boldsymbol{\theta}_g)$ is the conditional density of Y given $\mathbf{x}$ in the group Ω_g; the conditional densities belong to the same parametric family, indexed in $\boldsymbol{\theta}_g \in \boldsymbol{\Theta}$, $g = 1, \dots, G$.
- $\pi_g = p(\Omega_g)$ is the mixing weight of Ω_g, ($\pi_g > 0$ and $\sum_{g=1}^{G} \pi_g = 1$).
- $\boldsymbol{\psi} = (\pi_1, \dots, \pi_G, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_G)' \in \boldsymbol{\Psi}$ is the vector of all parameters.

Mixture of Linear Regression Models (MR) 2/3

In the simplest case, we assume that a set of G linear regression models characterized by the parameters $(\boldsymbol{\beta}_1, \sigma^2_{\varepsilon,1}), \ldots, (\boldsymbol{\beta}_G, \sigma^2_{\varepsilon,G})$ exist, and that for each observation pair $(Y_n, \mathbf{x}_n)$ the model

$$Y_n = \beta_{0g} + \boldsymbol{\beta}'_{1g}\mathbf{x}_n + \varepsilon_{n,g} \qquad g = 1, \ldots, G$$

where $\boldsymbol{\beta}_g = (\beta_{g0}, \boldsymbol{\beta}'_{1g})' = (\beta_{g0}, \beta_{g1}, \ldots, \beta_{gd})'$, $\mathbf{x}_n = (x_{n1}, \ldots, x_{nd})'$ and $\varepsilon_{n,g} \sim N(0, \sigma^2_{\varepsilon,g})$, holds with probability $\pi_g > 0$.

The quantities $\beta_1, \ldots, \beta_G$ as well as $\sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,G}$ are unknown parameters that need to be estimated from the data.

Mixture of Linear Regression Models (MR) 2/3

In the simplest case, we assume that a set of G linear regression models characterized by the parameters $(\boldsymbol{\beta}_1, \sigma^2_{\varepsilon,1}), \ldots, (\boldsymbol{\beta}_G, \sigma^2_{\varepsilon,G})$ exist, and that for each observation pair $(Y_n, \mathbf{x}_n)$ the model

$$Y_n = \beta_{0g} + \boldsymbol{\beta}'_{1g}\mathbf{x}_n + \varepsilon_{n,g} \qquad g = 1, \ldots, G$$

where $\boldsymbol{\beta}_g = (\beta_{g0}, \boldsymbol{\beta}'_{1g})' = (\beta_{g0}, \beta_{g1}, \ldots, \beta_{gd})'$, $\mathbf{x}_n = (x_{n1}, \ldots, x_{nd})'$ and $\varepsilon_{n,g} \sim N(0, \sigma^2_{\varepsilon,g})$, holds with probability $\pi_g > 0$.

The quantities $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G$ as well as $\sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,G}$ are unknown parameters that need to be estimated from the data.

Mixture of Linear Regression Models (MR) 3/3

Let us set $\widetilde{\mathbf{x}} = (1, \mathbf{x}')'$ and then $\boldsymbol{\beta}_g'\widetilde{\mathbf{x}}_n = \beta_{0g} + \boldsymbol{\beta}_{1g}'\mathbf{x}_n$.

Then, the conditional density of Y given the observation $\mathbf{x}_n$, can be written as

$$p(y|\mathbf{x}_n;\boldsymbol{\psi}) = \sum_{g=1}^{G} \phi(y;\boldsymbol{\beta}_g'\widetilde{\mathbf{x}}_n, \sigma^2_{\varepsilon,g})\,\pi_g, \qquad \pi_g \geq 0,\ \sum_{g=1}^{G}\pi_g = 1$$

where the vector ψ denotes the overall parameters of the model.

For each value of $\mathbf{x}_n$, the conditional distribution of Y_n given $\mathbf{x}_n$ is a mixture of univariate normal distributions with mean $\mu_{ng} = \boldsymbol{\beta}_g'\widetilde{\mathbf{x}}_n$ and variance $\sigma^2_{\varepsilon,g}$.

This model is also called *latent class regression* or *clusterwise regression*.

Remark

A finite mixture of regression models may be seen as an extension of a mixture of univariate normal distributions where the mean in the mixture distribution depends on explanatory variables.

On the other hand, a finite mixture of univariate normal distributions may be seen as that special case of finite mixtures of regression models where $\mu_{ng} = \beta_{g0}$ and $\boldsymbol{\beta}_g = \mathbf{0}$.

Mixture of Linear Regression Models (MR) 3/3

Let us set $\widetilde{\mathbf{x}} = (1, \mathbf{x}')'$ and then $\boldsymbol{\beta}'_g \widetilde{\mathbf{x}}_n = \beta_{0g} + \boldsymbol{\beta}'_{1g} \mathbf{x}_n$.

Then, the conditional density of Y given the observation $\mathbf{x}_n$, can be written as

$$p(y|\mathbf{x}_n; \boldsymbol{\psi}) = \sum_{g=1}^{G} \phi(y; \boldsymbol{\beta}'_g \widetilde{\mathbf{x}}_n, \sigma^2_{\varepsilon,g})\, \pi_g, \qquad \pi_g \geq 0, \; \sum_{g=1}^{G} \pi_g = 1$$

where the vector ψ denotes the overall parameters of the model.

For each value of $\mathbf{x}_n$, the conditional distribution of Y_n given $\mathbf{x}_n$ is a mixture of univariate normal distributions with mean $\mu_{ng} = \boldsymbol{\beta}'_g \widetilde{\mathbf{x}}_n$ and variance $\sigma^2_{\varepsilon,g}$.

This model is also called *latent class regression* or *clusterwise regression*.

Remark

A finite mixture of regression models may be seen as an extension of a mixture of univariate normal distributions where the mean in the mixture distribution depends on explanatory variables.

On the other hand, a finite mixture of univariate normal distributions may be seen as that special case of finite mixtures of regression models where $\mu_{ng} = \beta_{g0}$ and $\boldsymbol{\beta}_g = \mathbf{0}$.

Mixture of Linear Regression Models (MR) 3/3

Let us set $\widetilde{\mathbf{x}} = (1, \mathbf{x}')'$ and then $\boldsymbol{\beta}_g' \widetilde{\mathbf{x}}_n = \beta_{0g} + \boldsymbol{\beta}_{1g}' \mathbf{x}_n$.

Then, the conditional density of Y given the observation $\mathbf{x}_n$, can be written as

$$p(y|\mathbf{x}_n; \boldsymbol{\psi}) = \sum_{g=1}^{G} \phi(y; \boldsymbol{\beta}_g' \widetilde{\mathbf{x}}_n, \sigma^2_{\varepsilon,g})\, \pi_g, \qquad \pi_g \geq 0, \ \sum_{g=1}^{G} \pi_g = 1$$

where the vector ψ denotes the overall parameters of the model.

For each value of $\mathbf{x}_n$, the conditional distribution of Y_n given $\mathbf{x}_n$ is a mixture of univariate normal distributions with mean $\mu_{ng} = \boldsymbol{\beta}_g' \widetilde{\mathbf{x}}_n$ and variance $\sigma^2_{\varepsilon,g}$.

This model is also called *latent class regression* or *clusterwise regression*.

Remark

A finite mixture of regression models may be seen as an extension of a mixture of univariate normal distributions where the mean in the mixture distribution depends on explanatory variables.

On the other hand, a finite mixture of univariate normal distributions may be seen as that special case of finite mixtures of regression models where $\mu_{ng} = \beta_{g0}$ and $\boldsymbol{\beta}_g = \mathbf{0}$.

Mixture of Regressions with Concomitant (MRC)

Extension of MR is the *mixture of regressions with concomitant variables (MRC)*.

Mixtures of Regressions with Concomitant (MRC)

Consider a pair $(Y, \mathbf{X})$ of a response variable Y and covariates $\mathbf{X}$ defined on some population Ω. We say that the distribution of $Y|x$ is a mixture of regressions if has pdf given by

$$p(y|\mathbf{x}, \psi) = \sum_{g=1}^{G} f(y|\mathbf{x}; \boldsymbol{\theta}_g) p(\Omega_g|\mathbf{x}, \mathbf{w}) \, .$$

Mixture of Regressions with Concomitant (MRC)

The mixing weights $p(\Omega_g|\mathbf{x}, \mathbf{w})$ are now functions depending on $\mathbf{x}$ through some parameters $\mathbf{w}$, and $\boldsymbol{\psi} = (\pi_1, \ldots, \pi_G, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_G, \mathbf{w})$.

$p(\Omega_g|\mathbf{x}, \mathbf{w})$ is usually modeled by a multinomial logistic distribution with the first component as baseline, that is:

$$p(\Omega_g|\mathbf{x}, \mathbf{w}) = \frac{\exp(\mathbf{w}_g'\mathbf{x})}{\sum_{h=1}^{G} \exp(\mathbf{w}_h'\mathbf{x})}\ .$$

where $\mathbf{w}_g = (w_{g0}, w_{g1}, \ldots, w_{gd})' \in \mathbb{R}^{d+1}$ and $\mathbf{w} = (\mathbf{w}_1', \ldots, \mathbf{w}_G')' \in \mathbb{R}^{G(d+1)}$.

The EM algorithm for mixtures of simple regression 1/2

For simplicity, we present only the case of mixtures of simple regression models:

$$p(x;\boldsymbol{\psi}) = \pi_1\phi(y;\beta_{10}+\beta_{11}x,\sigma_1^2)+\cdots+\pi_G\phi(y;\beta_{G0}+\beta_{G1}x,\sigma_G^2)$$

Let $\mathbf{x}=(x_1,\ldots,x_N)'$ be a sample taken from $p(x;\boldsymbol{\psi})$, the complete log-likelihood functionis given by

$$\mathcal{L}_c(\boldsymbol{\psi}) = \sum_{n=1}^{N}\sum_{g=1}^{G} z_{ng}\log\pi_g + \sum_{n=1}^{N}\sum_{g=1}^{G} z_{ng}\log\phi(y_n;\beta_{g0}+\beta_{g1}x_n,\sigma_g^2)\,,$$

where the vectors $\mathbf{z}_1,\ldots,\mathbf{z}_N$ are the missing data which are realizations of the random vectors $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$, where $Z_{ng}=(\mathbf{Z}_n)_g$, is defined as $Z_{ng}=1$ if x_n comes from Ω_g and $Z_{ng}=0$ otherwise, for $n=1,\ldots,N$ and $g=1,\ldots,G$.

The EM algorithm for mixtures of simple regression 2/2

We get the following estimates at step r:

$$\tau_{ng}^{(r+1)} = \tau_{ng}(\mathbf{x}; \boldsymbol{\psi}^{(r)}) = \frac{\pi_g^{(r)} \phi\left(x_n; \beta_{g0}^{(r)}, \beta_{g1}^{(r)}, \sigma_g^{2(r)}\right)}{p(x_n; \boldsymbol{\psi}^{(r)})}$$

$$\pi_g^{(r+1)} = \frac{1}{N} \sum_{n=1}^{N} \tau_{ng}^{(r+1)}$$

$$\beta_{g1}^{(r+1)} = \frac{\sum_{n=1}^{N} \tau_{ng}^{(r+1)} (y_n - \bar{y}_g)(x_n - \bar{x}_g)}{\sum_{n=1}^{N} \tau_{ng}^{(r+1)} (x_n - \bar{x}_g)^2}$$

$$\beta_{g0}^{(r+1)} = \bar{y}_g - \beta_{g1}^{(r+1)} \bar{x}_g$$

$$\left(\sigma_g^2\right)^{(r+1)} = \frac{\sum_{n=1}^{N} \tau_{ng}^{(r+1)} \left(y_n - \beta_{g0}^{(r+1)} - \beta_{g1}^{(r+1)} x_n\right)^2}{\sum_{n=1}^{N} \tau_{ng}^{(r+1)}}$$

Parameter estimation and classification

Assume we are provided with a set of N independent observation pairs $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ drawn from a mixture of regressions (MR/MRC). Then:

1. for fixed G, estimate model parameters, usually according to the maximum likelihood approach, using the EM algorithm;
2. if G is unknown, then
    1. repeat step 1 for different number G of groups;
    2. select G according to model selection criteria like BIC or ICL;
3. based on the estimate $\widehat{\psi}$, compute the posterior probability $\tau_g(\mathbf{x}_n, y_n; \widehat{\psi})$ that the nth unit $(y_n, \mathbf{x}_n)$ belong to the gth group Ω_g:

$$\text{MR}: \quad \tau_g(\mathbf{x}_n, y_n|\widehat{\psi}) = \frac{f(y_n|\mathbf{x}_n, \widehat{\boldsymbol{\theta}}_g)\widehat{\pi}_g}{p(y_n|\mathbf{x}, \widehat{\psi})} = \frac{f(y_n|\mathbf{x}_n; \widehat{\boldsymbol{\theta}}_g)\widehat{\pi}_g}{\sum_{h=1}^{G} f(y_n|\mathbf{x}_n, \widehat{\boldsymbol{\theta}}_h)\widehat{\pi}_h}$$

$$\text{MRC}: \quad \tau_g(\mathbf{x}_n, y_n|\widehat{\psi}) = \frac{f(y_n|\mathbf{x}_n, \widehat{\boldsymbol{\theta}}_g)p(\Omega_g|\mathbf{x}_n, \widehat{\mathbf{w}})}{p(y|\mathbf{x}; \widehat{\psi})} = \frac{f(y_n|\mathbf{x}_n, \widehat{\boldsymbol{\theta}}_g)\exp(\mathbf{w}_g'\mathbf{x}_n)}{\sum_{h=1}^{G} f(y_n|\mathbf{x}_n, \widehat{\boldsymbol{\theta}}_h)\exp(\mathbf{w}_h'\mathbf{x}_n)}.$$

and classify units into groups according to the maximum a posteriori probability (MAP) criterion.

Model selection

The problem

How to estimate the number G of groups?

- For model selection, usually BIC and ICL.

Criterion	*Definition*
BIC	$= -2\mathcal{L}(\psi) + K \log N$
ICL	$\approx$ BIC + $\sum_n 1_{\hat{z}_{ig}} \ln \hat{z}_{ig}$

are usually adopted, where K is the number of parameter in the model.

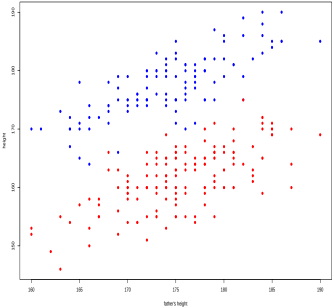
Note

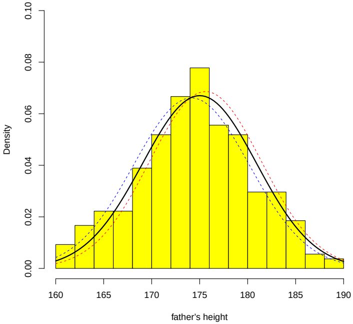
There may be no particular reasons for choosing a single best model over the other ones. On the contrary, it makes more sense to "deselect" models that are obviously poor, maintaining a subset for further considerations.

Example A: Student data (cont'd) 2/4

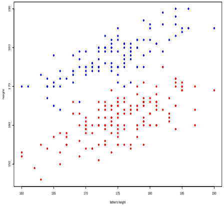
The relationship between *height of the respondent* and *height of respondent's father* has been modeled according to both finite mixture of regression (MR) and finite mixture of regression with concomitant (MRC).

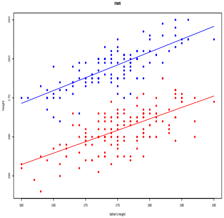
The histogram of the X-variable *height of respondent's father* does not show any cluster structure along the X-variable.

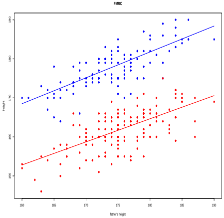




Example A: Student data (cont'd) 3/4







The two models give in practice the same results

Example A: Student data (cont'd) 4/4

	MR	MRC
Confusion Matrices	see below	see below
Misclassification Error	2.59%	2.22%
Adjusted Rand Index	0.8986	0.9127

MR

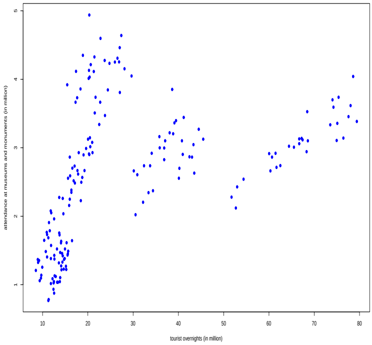
Actual	Predicted M	Predicted F
M	112	7
F	0	151

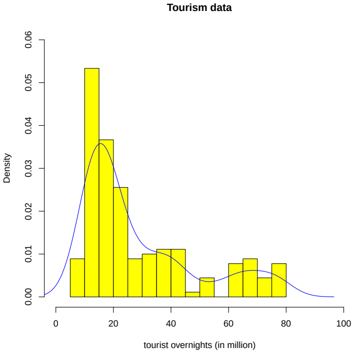
MRC

Actual	Predicted M	Predicted F
M	113	6
F	0	151

Example B: Tourism data (cont'd - 2/4)

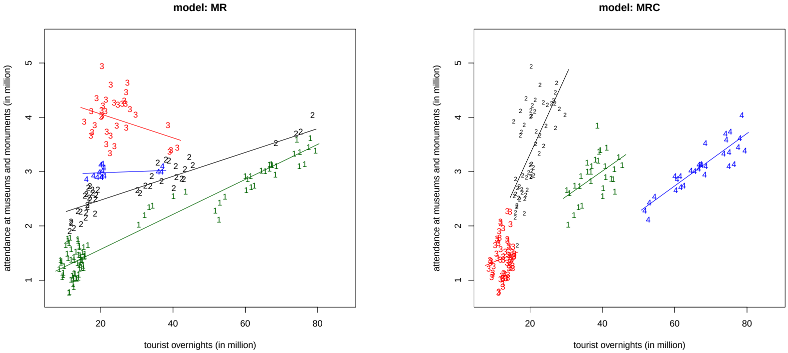
Cluster structure with respect to the covariate:





The data have been modeled without considering the time information (month labels)

Example B: Tourism data (cont'd - 3/4)



- choice: $G = 4$ for economic reasons.

Example B: Tourism data (cont'd - 4/4)

Model: *Mixture of Regressions with Concomitant (FMRC)*.

Group	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	15	0	0	15	0	0	0
2	0	**1**	12	15	15	0	0	0	0	15	0	0
3	15	14	**3**	0	0	0	0	0	0	0	15	15
4	0	0	0	0	0	0	15	15	0	0	0	0

The four groups identify almost perfectly the units from time point of view.

Group 1 : *units in June and September (early and late summer);*

Group 2 : *units in March, April, May and October (spring and early autumn);*

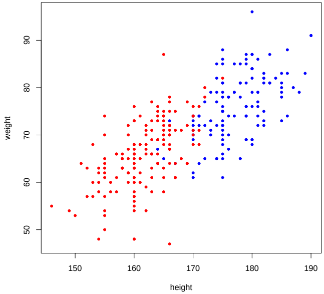
Group 3 : *units from November to February (late autumn and winter);*

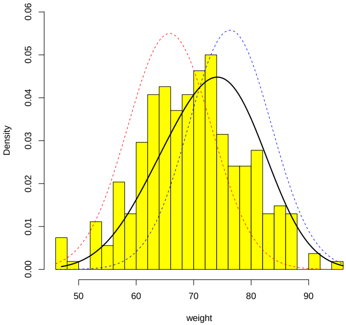
Group 4 : *units in July and August (summer);*

misclassification error: $\delta = 2.22\%$

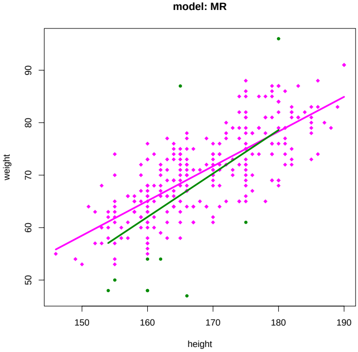
Example C: Student data 2 (cont'd) 1/2

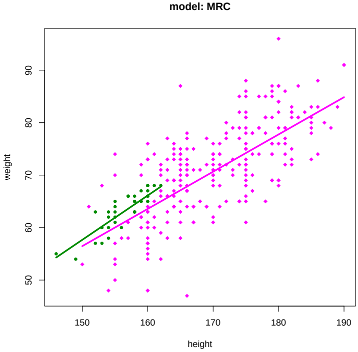
In *Student Data* consider the relationship between weight and height.
Two groups: males and females (blue=males, red=females) .





Example C: Student data 2 (cont'd) 2/2





$\delta = 46.30\%$

ARI = 0.0072

$\delta = 42.96\%$

ARI = 0.0105

Question

Can we do better?

Mixture of generalized linear models 1/3

More in general, consider mixture models of the distributions of $Y|\mathbf{x}, \Omega_g$ taking the form

$$p(y|\mathbf{x};\boldsymbol{\psi}) = \sum_{g=1}^{G} f(y|\mathbf{x};\boldsymbol{\theta}_g)\,\pi_g, \quad \text{with } \pi_g \geq 0, \sum_{g=1}^{G} \pi_g = 1$$

where Y is a dependent variable taking values in some space $\mathcal{Y} \subseteq \mathbb{R}$ with conditional density p, $\mathbf{x}$ is a vector of independent variables, π_g is the prior probability of component g, $\boldsymbol{\theta}_g \in \boldsymbol{\Theta}$ is the component specific parameter for the density f and $\boldsymbol{\psi} = (\pi_1, \ldots, \pi_G, \boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_G')' \in \boldsymbol{\Psi}$ is the vector of all parameters.

In order to deal with various response types, we assume that, for each mixture component, the conditional distributions $f(y|\mathbf{x};\boldsymbol{\theta}_g)$ belong to the exponential family.

This yields the *mixture of generalized linear models* known as *GLIMMIX* models in the marketing literature:

$$p(y|\mathbf{x};\boldsymbol{\psi}) = \sum_{g=1}^{G} f(y|\mathbf{x};\boldsymbol{\beta}_g, \lambda_g)\,\pi_g, \qquad \mathbf{x} \in \mathbb{R}^{d+1}, y \in \mathcal{Y}\,.$$

Mixture of generalized linear models 2/3

The binomial response.

Assume that Y takes values in $\mathcal{Y} = \{0, 1, \ldots, M\}$, for some $M \in \mathbb{N}$, and that the probability mass function of $Y|\mathbf{x}, \Omega_g$ is binomial with parameters $\left(M, \mu_g(\mathbf{x}; \boldsymbol{\beta}_g)/M\right)$, that is $Y|\mathbf{x}, \Omega_g \sim \text{Bin}\left(M, \mu_g\left(\mathbf{x}; \boldsymbol{\beta}_g\right)/M\right)$. In this case

$$f\left(y|\mathbf{x}; M, \boldsymbol{\beta}_g\right) = \binom{M}{y} \left[\frac{\mu_g(\mathbf{x}; \boldsymbol{\beta}_g)}{M}\right]^y \left[1 - \frac{\mu_g(\mathbf{x}; \boldsymbol{\beta}_g)}{M}\right]^{M-y},$$

where

$$\mu_g(\mathbf{x}; \boldsymbol{\beta}_g) = M \frac{\exp(\boldsymbol{\beta}_g' \widetilde{\mathbf{x}})}{1 + \exp(\boldsymbol{\beta}_g' \widetilde{\mathbf{x}})}.$$

Mixture of generalized linear models 3/3

The Poisson response.

Assume that Y takes values in $\mathcal{Y} = \mathbb{N}$ and that the probability mass function of $Y|\mathbf{x}, \Omega_g$ is Poisson with parameter μ_g, that is $Y|\mathbf{x}, \Omega_g \sim \text{Poi}\left(\mu_g\left(\mathbf{x}; \boldsymbol{\beta}_g\right)\right)$. In this case

$$f(y|\mathbf{x}; \boldsymbol{\beta}_g) = \exp\left[-\mu_g(\mathbf{x}; \boldsymbol{\beta}_g)\right] \frac{\left[\mu_g\left(\mathbf{x}; \boldsymbol{\beta}_g\right)\right]^y}{y!},$$

where

$$\mu_g(\mathbf{x}; \boldsymbol{\beta}_g) = \exp(\boldsymbol{\beta}_g' \widetilde{\mathbf{x}}).$$

Algorithms and software

flexmix: Flexible Mixture Modeling

A general framework for finite mixtures of regression models using the EM algorithm is implemented. The E-step and all data handling are provided, while the M-step can be supplied by the user to easily define new models. Existing drivers implement mixtures of standard linear models, generalized linear models and model-based clustering.

Version:	2.3-17
Depends:	R (≥ 2.15.0), lattice
Imports:	graphics, grid, grDevices, methods, modeltools (≥ 0.2-16), nnet, stats, stats4, utils
Suggests:	actuar, codetools, diptest, Ecdat, ellipse, gclus, glmnet, lme4 (≥ 1.1), MASS, mgcv (≥ 1.8-0), mlbench, multcomp, mvtnorm, SuppDists, survival
Published:	2020-10-12
Author:	Bettina Gruen [aut, cre], Friedrich Leisch [aut], Deepayan Sarkar [ctb], Frederic Mortier [ctb], Nicolas Picard [ctb]
Maintainer:	Bettina Gruen <Bettina.Gruen at R-project.org>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
NeedsCompilation:	no
Citation:	flexmix citation info
Materials:	NEWS
In views:	Cluster, Environmetrics, Psychometrics
CRAN checks:	flexmix results

Labo activity with R

Lab_activity_MM_2.R