Semiparametric Mixture Regression with Unspecified Error Distributions

Weixin Yao[1]
(Joint work with Yanyuan Ma[2], Shaoli Wang[3] and Lin Xu[4])

[1]University of California, Riverside, [2]The Pennsylvania State University & [3]Shanghai University of Finance and Economics, [4]Zhejiang University of Finance and Economics

Presented at

Harnessing the power of latent structure models and modern Big Data learning

Outline

1. Introduction of mixtures of regressions
   - Real examples
   - Model setting
   - Estimating algorithm
2. New model: Mixtures of regressions with unspecified error density
   - Model setting
   - Identifiable result
   - Estimating algorithm
   - Asymptotic properties
   - Simulation study
3. Some extensions and future work

Model setting of conventional mixtures of regressions

- Latent class variable $\mathcal{Z}$: $P(\mathcal{Z} = j) = \pi_j$ for $j = 1, 2, \cdots, m$.
- Given $\mathcal{Z} = j$,
$$Y = \mathbf{X}^T \beta_j + \epsilon_j.$$
- $\epsilon_j \sim N(0, \sigma_j^2)$.
- $\mathbf{X}$, $\mathcal{Z}$, and ϵ are jointly independent.

$\Rightarrow$ Without observing $\mathcal{Z}$,

$$Y|\mathbf{X} = \mathbf{x} \sim \sum_{j=1}^{m} \pi_j N(\mathbf{x}^T \beta_j, \sigma_j^2).$$

Likelihood function

Given observations $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, the log-likelihood function is

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \log\left\{\pi_j \phi(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)\right\},$$

where $\phi(t; \mu, \sigma^2)$ is the density of $N(\mu, \sigma^2)$.

$\Rightarrow$ No explicit solution for MLE.

Estimating algorithm–EM algorithm

- Let $\boldsymbol{\theta} = (\pi_1, \sigma_1, \boldsymbol{\beta}_1, \ldots, \pi_m, \sigma_m, \boldsymbol{\beta}_m)$.
- Starting with the initial parameter $\boldsymbol{\theta}^{(0)}$, at $(k+1)$th step:

  E step: find the classification probabilities

$$\begin{aligned} p_{ij}^{(k+1)} &= P(\mathcal{Z}_i = j \mid \boldsymbol{\theta}^{(k)}) \\ &= \frac{\pi_j^{(k)} \phi(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)}, \sigma_j^{2(k)})}{\sum_{j=1}^m \pi_j^{(k)} \phi(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)}, \sigma_j^{2(k)})}. \end{aligned}$$

Estimating algorithm–EM algorithm

M step: Update $\boldsymbol{\theta}$

$$
\begin{aligned}
\boldsymbol{\beta}_j^{(k+1)} &= \arg\min_{\boldsymbol{\beta}_j} \sum_{i=1}^{n} p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2, \\
\pi_j^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} p_{ij}^{(k+1)}, \\
\sigma_j^{2(k+1)} &= \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k+1)} \right)^2}{\sum_{i=1}^{n} p_{ij}^{(k+1)}}.
\end{aligned}
$$

Comments:

- For linear regression, LSE does not require normal assumption of error.
- For conventional mixtures of linear regressions, the MLE is invalid if the normal assumption is violated and the normal density is explicitly used in E step of EM algorithm.

$\Rightarrow$ How to relax the strong normal assumption of error density?

Model setting for unknown error density

- Latent class variable $\mathcal{Z}$: $P(\mathcal{Z} = j) = \pi_j$ for $j = 1, 2, \cdots, m$.
- Given $\mathcal{Z} = j$, $Y = \mathbf{X}^T\beta_j + \epsilon_j$.
- $\mathbf{X}$, $\mathcal{Z}$, and ϵ are jointly independent.
- $\mathrm{E}(\epsilon_j \mid \mathbf{X}) = 0$ and ϵ_j has the density

$$g_j(t) = \frac{1}{\sigma_j} g(t/\sigma_j),$$

where $g(\cdot)$ is unknown with mean 0 and variance 1.

1

[1] Ma, Y., Wang, S., Xu, L., & Yao, W. (2021). Semiparametric mixture regression with unspecified error distributions. *Test*, 30(2), 429-444.

Identifiable result

- New model density:

$$f(y|\mathbf{x},\boldsymbol{\theta}) = \sum_{j=1}^{m} \pi_j g(\epsilon_j/\sigma_j)/\sigma_j,$$

  where $\epsilon_j = y - \mathbf{x}^T\boldsymbol{\beta}_j$ and $\boldsymbol{\theta} = (\pi_1, \sigma_1, \boldsymbol{\beta}_1, \ldots, \pi_m, \sigma_m, \boldsymbol{\beta}_m, g)$.

- Let $\boldsymbol{\omega} = (\boldsymbol{\omega}(1), \ldots, \boldsymbol{\omega}(m))$ be any permutation of $(1, \ldots, m)$.
- $\boldsymbol{\theta}^{\boldsymbol{\omega}} = (\pi_{\boldsymbol{\omega}(1)}, \sigma_{\boldsymbol{\omega}(1)}, \boldsymbol{\beta}_{\boldsymbol{\omega}(1)}, \ldots, \pi_{\boldsymbol{\omega}(m)}, \sigma_{\boldsymbol{\omega}(m)}, \boldsymbol{\beta}_{\boldsymbol{\omega}(m)}, g)$.

$\Rightarrow f(y|\mathbf{x},\boldsymbol{\theta}) = f(y|\mathbf{x},\boldsymbol{\theta}^{\boldsymbol{\omega}})$, for any $\boldsymbol{\omega}$.

Identifiable result–definition

Reminder of model density:

$$f(y|\mathbf{x},\boldsymbol{\theta}) = \sum_{j=1}^{m} \frac{\pi_j}{\sigma_j} g\left(\frac{y-\mathbf{x}^T\boldsymbol{\beta}_j}{\sigma_j}\right).$$

Definition

We say the mixtures of regressions model $f(y|\mathbf{x},\boldsymbol{\theta})$ is identifiable for the parameter $\boldsymbol{\theta}$ if $f(y|\mathbf{x},\boldsymbol{\theta}_1) = f(y|\mathbf{x},\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2^{\boldsymbol{\omega}}$ for some permutation $\boldsymbol{\omega}$.

Identifiable result

Let $\mathbf{x} = (1, \mathbf{x}_s^T)^T, \boldsymbol{\beta}_j = (\beta_0, \boldsymbol{\beta}_{sj})^T$ with $\boldsymbol{\beta}_{sj} = (\beta_{1j}, \ldots, \beta_{pj})^T$.

Theorem

Suppose that $\pi_j > 0$ and $\boldsymbol{\beta}_{sj}$'s are distinct vectors in $\mathbb{R}^p$. If the domain of $\mathbf{x}$ contains an open set in $\mathbb{R}^p$, then the semiparametric regression model $f(y|\mathbf{x}, \boldsymbol{\theta})$ is identifiable.

Remark: a) The identifiable result also holds for the more general model:

$$f(y|\mathbf{x}, \boldsymbol{\theta}) = \sum_{j=1}^{m} \pi_j g_j(y - \mathbf{x}^T \boldsymbol{\beta}_j).$$

b) When $\mathbf{x} = 1$, the identifiable result requires $g_j(\cdot) = g(\cdot)$ and $g(\cdot)$ is symmetric about 0.

(Bordes, et al., 2006 and Hunter, et al., 2007)

Estimating algorithm

- Kernel DEnsity based EM type algorithm (KDEEM)

  E step: Calculate the classification probabilities,

$$\begin{aligned} p_{ij}^{(k+1)} &= P(\mathcal{Z}_i = j) \\ &= \frac{\pi_j^{(k)} g^{(k)}(r_{ij}^{(k)})/\sigma_j^{(k)}}{\sum_{j=1}^{m} \pi_j^{(k)} g^{(k)}(r_{ij}^{(k)})/\sigma_j^{(k)}}, \end{aligned}$$

  where $r_{ij}^{(k)} = e_{ij}^{(k)}/\sigma_j^{(k)}$ and $e_{ij}^{(k)} = y_i - x_i^T \beta_j^{(k)}$.

Estimating algorithm

M step:

$$\pi_j^{(k+1)} = \frac{\sum_{i=1}^n p_{ij}^{(k+1)}}{n},$$

$$(\beta_j^{(k+1)T}, \sigma_j^{(k+1)})^T = \arg\max_{\beta_j,\sigma_j} \sum_{i=1}^n p_{ij}^{(k+1)} \log[g^{(k)}\{(y_i - \mathbf{x}_i^T\beta_j)/\sigma_j\}/\sigma_j],$$

$$g^{(k+1)}(t) = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^m p_{ij}^{(k+1)} K_h(r_{ij}^{(k+1)} - t),$$

where $K_h(t) = h^{-1}K(t/h)$ and $K(t)$ is a kernel function, such as Gaussian or Epanechnikov kernel.

Estimating algorithm–some special cases

Some special cases:

- Same/homogeneous component densities, i.e., $\sigma_j = \sigma$ for all js (Hunter and Young, 2012). Denote the resulting estimate by KDEEM.H.

- Least squares criterion to update $\boldsymbol{\beta}$ (Hunter and Young, 2012)

$$\boldsymbol{\beta}_j^{(k+1)} = \arg\min_{\boldsymbol{\beta}_j} \sum_{i=1}^{n} p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2.$$

  Denote the resulting estimate by KDEEM.LSE.

Asymptotic properties

For KDEEM, KDEEM.H, and KDEEM.LSE, we have the following asymptotic results.

Theorem

Under some regularity conditions,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, V),$$

in distribution when $n \to \infty$.
In addition, $\hat{g}(t) - g_0(t) = O_p\{h^2 + (nh)^{-1/2}\}$.

Example 1

- Generate i.i.d. data $\{(x_i, y_i), i = 1, \ldots, n\}$ from the model

$$Y = \begin{cases} -3 + 3X + \epsilon_1, & \text{if } \mathcal{Z} = 1; \\ 3 - 3X + \epsilon_2, & \text{if } \mathcal{Z} = 2. \end{cases}$$

- $X \sim U(0, 1)$.
- $P(\mathcal{Z} = 1) = 0.5$.
- $\epsilon_2 \sim 0.5\epsilon_1$.

Example 1– error densities considered

- Consider the following cases for ϵ_1:

Case I: $\epsilon_1 \sim N(0,1)$

Case II: $\epsilon_1 \sim U(-3,3)$

Case III: Two-modal pdf

$$\epsilon_1 \sim 0.5N(-1.5, 0.5^2) + 0.5N(1.5, 0.5^2)$$

Case IV: Skewed pdf

$$\epsilon_1 \sim 0.5N(-0.7, 0.5^2) + 0.5N(0.7, 1.5^2)$$

Case V: $\epsilon_1 \sim \Lambda(0, 1^2)$

Case VI: $\epsilon_1 \sim Gamma(2, 0.5)$

Case VII: $\epsilon_1 \sim Rayleigh(3)$

Example 1–continue

- Consider the following four estimation methods:

  MLEEM: Normal assumption based MLE via the EM algorithm

  KDEEM: The proposed Semiparametric EM algorithm with unspecified error density

  KDEEM.H: A special case of KDEEM assuming homogeneous component variance

  KDEEM.LSE: A special case of KDEEM.H using least squares to update component regression parameters

Case I-IV: Relative efficiency between different estimates and MLEEM when n=250

Error distributions		MLEEM	KDEEM	KDEEM.H	KDEEM.LSE
Case I	$\beta_{1,0}$	1	0.76	0.62	0.74
$\epsilon_1 \sim N(0,1)$	$\beta_{1,1}$	1	0.84	0.57	0.65
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	0.77	0.45	0.36
	$\beta_{1,1}$	1	0.81	0.74	0.70
Case II	$\beta_{1,0}$	1	1.46	0.62	0.23
$\epsilon_1 \sim U(-3,3)$	$\beta_{1,1}$	1	3.01	0.70	0.51
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	1.57	0.36	0.16
	$\beta_{1,1}$	1	3.44	1.07	0.34
Case III	$\beta_{1,0}$	1	2.65	0.67	0.58
$\epsilon_1 \sim 0.5N(-1.5,0.5^2)+0.5N(1.5,0.5^2)$	$\beta_{1,1}$	1	12.93	1.43	0.37
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	4.27	1.76	0.24
	$\beta_{1,1}$	1	20.15	15.55	0.66
Case IV	$\beta_{1,0}$	1	1.12	0.93	1.29
$\epsilon_1 \sim 0.5N(-1,0.5^2)+0.5N(1,1.5^2)$	$\beta_{1,1}$	1	3.41	2.01	1.62
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	4.23	3.15	1.12
	$\beta_{1,1}$	1	7.53	7.78	1.21

Case V-VII: Relative efficiency between different estimates and MLEEM when n=250

Error distributions		MLEEM	KDEEM	KDEEM.H	KDEEM.LSE
Case V	$\beta_{1,0}$	1	1.16	1.57	4.15
$\epsilon_1 \sim \Lambda(0, 1^2)$	$\beta_{1,1}$	1	8.83	4.93	0.85
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	3.76	5.93	3.94
	$\beta_{1,1}$	1	20.34	25.81	1.44
Case VI	$\beta_{1,0}$	1	1.35	0.79	1.12
$\epsilon_1 \sim Gamma(2, 0.5)$	$\beta_{1,1}$	1	2.82	1.68	1.08
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	0.96	0.24	1.00
	$\beta_{1,1}$	1	2.11	1.37	1.12
Case VII	$\beta_{1,0}$	1	1.27	0.95	0.64
$\epsilon_1 \sim Rayleigh(3)$	$\beta_{1,1}$	1	1.47	0.98	0.80
$\epsilon_2 \sim 0.5\epsilon_1$	$\beta_{1,0}$	1	1.19	1.05	0.59
	$\beta_{1,1}$	1	1.26	1.25	1.13

Equine Infectious Anemia Virus data

- 8 horses were inoculated with Equine Infectious Anemia Virus (EIAV) infectious clone and 5 horses were inoculated vaccine strain.
- 45 observations were obtained from 8 mixed-gender horses.
- After the 15 days immunization period, five horses inoculated vaccine (39 observations, ID 1 to 39) were normal. Three horses who were not vaccinated (6 observations, ID 40 to 45) had fever and two of them died at the end of the experiment.
- Response: log value of viral loads which measure the immune ability of the infected horses
- Predictors: SLFN11, viperin, Trim5a, A3G, IFITM, SAMHD1 and Tetherin, which can be used in immunodiffusion assay to confirm wether an animal was protected.

Estimation results for equine infectious anemia virus data

Covariate	MLEEM		KDEEM	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Intercept	7.72	17.00	8.60	22.47
SLFN11	-0.06	-0.18	-0.08	0.51
Viperin	0.77	-0.19	-0.33	0.13
Trim5a	-0.44	-1.17	0.76	-6.06
A3G	-0.23	-1.16	-0.14	-2.36
IFITM	-0.54	-1.47	-0.56	-1.62
SAMHD1	0.11	-4.48	0.31	-1.86
Tetherin	0.12	1.81	-0.31	2.48
CCP	62.22%		100%	

CCP: correct classification percentages

Equine infectious anemia virus data

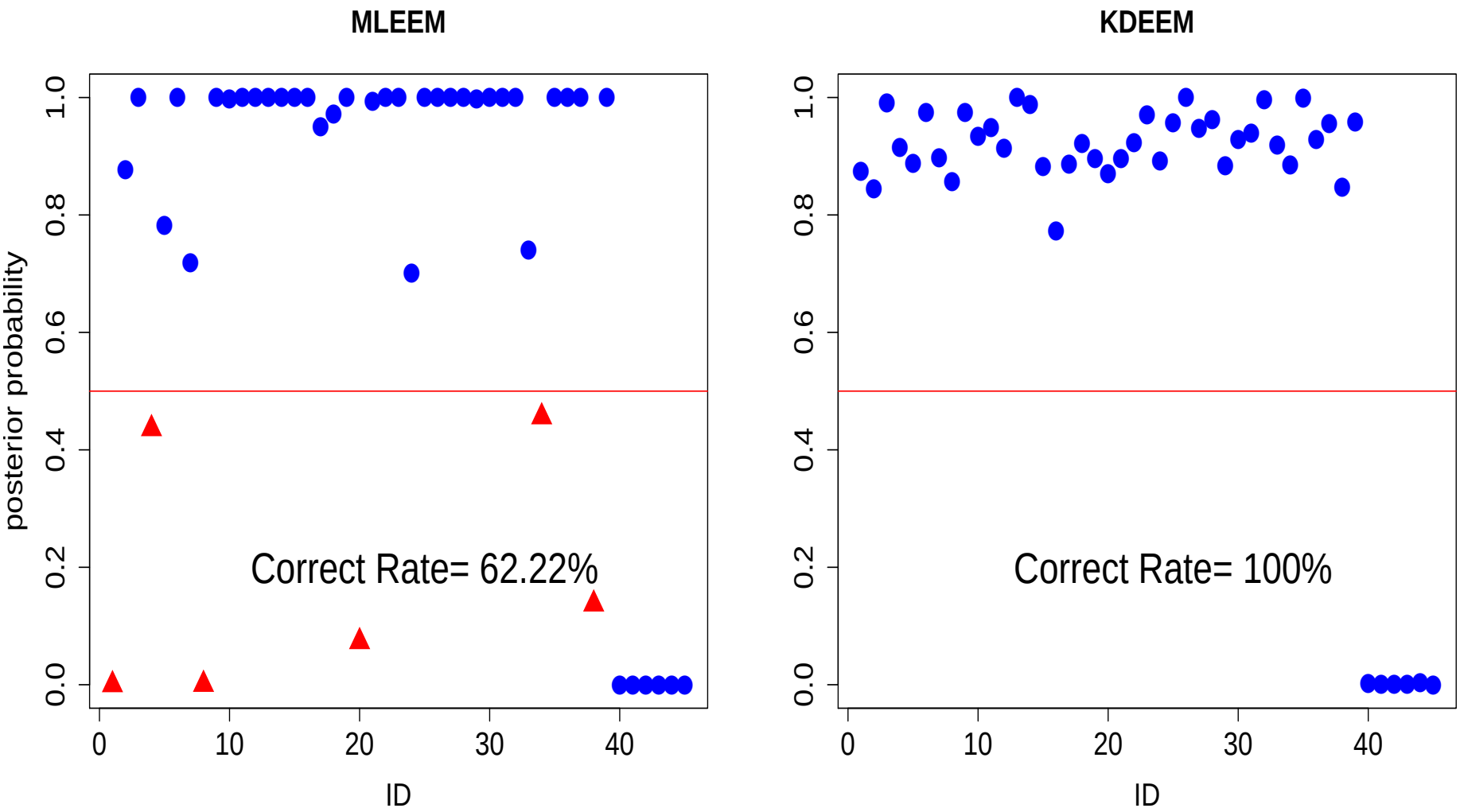


Figure: Comparison of classification results

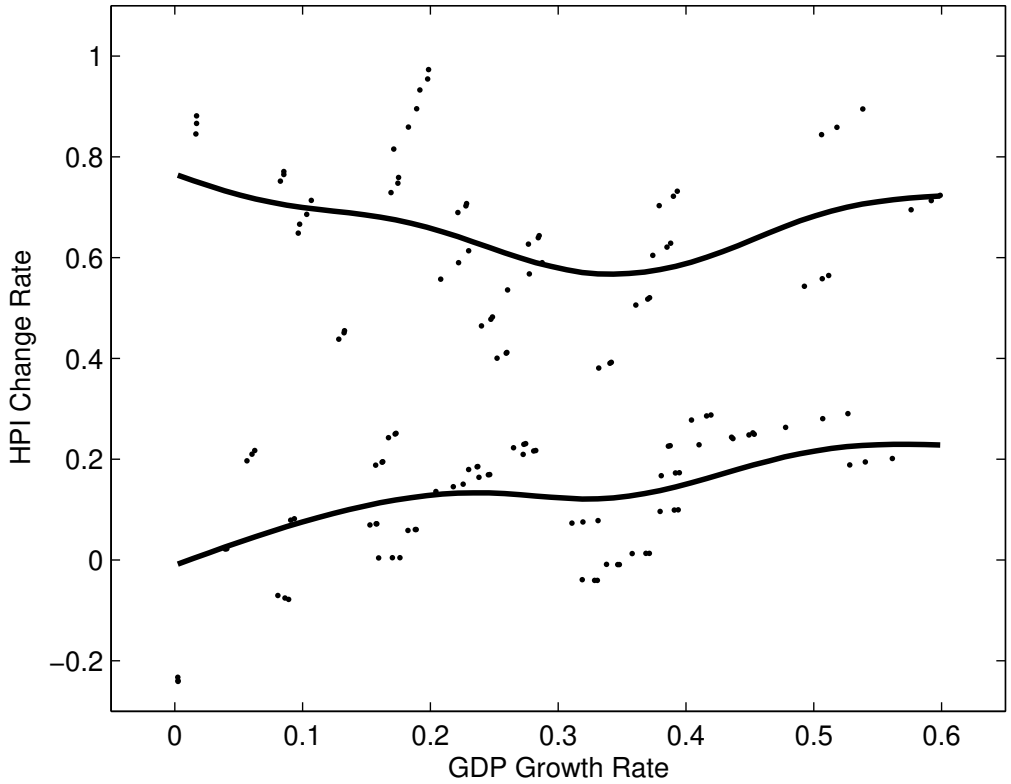
A motivating example–US house price index data

- Response: House price index (HPI) change ate (1990-2002)

  Predictor: GDP growth rate

- Two different patterns in different macroeconomic cycles.

- Two patterns are not linear.



Extension to mixtures of nonparametric regressions

Model setting:

- Latent class variable $\mathcal{Z}$: $P(\mathcal{Z} = j \mid x) = \pi_j(x)$ for $j = 1, 2, \cdots, m$.
- Given $\mathcal{Z} = j$ and $X = x$,

$$Y = m_j(x) + \epsilon_j,$$

where $m_j(x)$ is an unknown smoothing function.

Extensions

- Let $g_j(\epsilon \mid x)$ be the density of ϵ given X, such that

$$g_j(\epsilon \mid x) = \frac{1}{\sigma_j(x)} g(\epsilon/\sigma_j(x)),$$

where $g(\cdot)$ is unknown with mean 0 and variance 1.

$\Rightarrow$ Given $X = x$,

$$f(y \mid x) = \sum_{j=1}^{m} \frac{\pi_j(x)}{\sigma_j(x)} g\left(\frac{y - m_j(x)}{\sigma_j(x)}\right).$$

Extension to mixtures of nonparametric regressions

- Identifiable result follows directly from linear regression
- If $\mathcal{Z}$ is known, Nadaraya-Watson estimator gives

$$
\begin{aligned}
m_j(x) &= \frac{\sum_{i=1}^n K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j) y_i}{\sum_{i=1}^n K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j)}, \\
\sigma_j^2(x) &= \frac{\sum_{i=1}^n K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j)(y_i - m_j(x))^2}{\sum_{i=1}^n K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j)}, \\
\pi_j(x) &= \frac{\sum_{i=1}^n K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j)}{\sum_{i=1}^n \sum_{j-1}^m K(x_i - x)\mathrm{I}(\mathcal{Z}_i = j)}.
\end{aligned}
$$

Estimating algorithm

- Kernel regression based EM type algorithm:

  E step:

$$\begin{aligned} p_{ij}^{(k+1)} &= P(\mathcal{Z}_i = j \mid \boldsymbol{\theta}^{(k)}) \\ &= \frac{\pi_j^{(k)} g^{(k)}(r_{ij}^{(k)})/\sigma_j^{(k)}}{\sum_{j=1}^{m} \pi_j^{(k)} g^{(k)}(r_{ij}^{(k)})/\sigma_j^{(k)}}, \end{aligned}$$

  where $e_{ij}^{(k)} = y_i - m_j^{(k)}(x_i)$ and $r_{ij}^{(k)} = e_{ij}^{(k)}/\sigma_j^{(k)}(x_i)$.

Estimating algorithm

M step:

$$
\begin{aligned}
m_j^{(k+1)}(x) &= \frac{\sum_{i=1}^n p_{ij}^{(k+1)} K_h(x_i - x) y_i}{\sum_{i=1}^n p_{ij}^{(k+1)} K_h(x_i - x)}, \\
\pi_j^{(k+1)}(x) &= \frac{\sum_{i=1}^n p_{ij}^{(k+1)} K_h(x_i - x)}{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} K_h(x_i - x)}, \\
\sigma_j^{2(k+1)}(x) &= \frac{\sum_{i=1}^n p_{ij}^{(k+1)} K_h(x_i - x)(e_{ij}^{(k+1)})^2}{\sum_{i=1}^n p_{ij}^{(k+1)} K_h(x_i - x)}, \\
g^{(k+1)}(t) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} K_h(r_{ij}^{(k+1)} - t).
\end{aligned}
$$

Estimating algorithm–some special cases

Some special cases:

- If $\sigma_j(x) = \sigma(x)$ for all js,

$$\sigma^{2(k+1)}(x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} K_h(x_i - x)(e_{ij}^{(k+1)})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} K_h(x_i - x)}.$$

- If $\pi_j(x) = \pi_j$ for all js and x,

$$\pi_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} p_{ij}^{(k+1)}.$$

References

- Benaglia, T., Chauveau, D., and Hunter D. R. (2009). An EM-like algorithm for semi- and non-parametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, 18, 505-526.
- Bordes, L., Chauveau, D., and Vandekerkhove, P. (2007). An EM algorithm for a semiparametric mixture model. *Computational Statistics and Data Analysis*, 51, 5429-5443.
- Bordes, L., Mottelet, S., and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics*, 34, 1204-1232.
- Hunter, D., Wang, S., and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *Annals of Statistics*, 35, 224-251.

References

- Hunter, D. R. and Young, D. S. (2012). Semiparametric mixtures of regressions. *Journal of Nonparametric Statistics*, 24:19-38.
- Cohen, E. (1984). Some effects of inharmonic partials on interval perception. *Music Perception*, 1, 323-349.
- Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society*, Ser. B, 60, 271-294.
- Ma, Y., Wang, S., Xu, L., & **Yao, W.** (2021). Semiparametric mixture regression with unspecified error distributions. *Test*, 30(2), 429-444.


THANK
YOU
Profile-Comments.com