Fuzzy Regression Models

Fuzzy Clustering and Switching Regression

Olaf Wolkenhauer

Control Systems Centre

UMIST

o.wolkenhauer@umist.ac.uk

www.csc.umist.ac.uk/people/wolkenhauer.htm

Contents

1. Switching Regression

Instead of assuming that a single regression model, can account for the data $\mathbf{m}_j = (\mathbf{x}_j, y_j)$, a *switching regression model* is specified by

$$y = f_i(\mathbf{x};\boldsymbol{\theta}_i) + \varepsilon_i \ , \qquad 1 \le i \le c \tag{1}$$

In a statistical framework the optimal estimate of $\boldsymbol{\theta}$ depends on assumptions made about the distribution of random vectors ε_i. Commonly, the ε_i are assumed to be independently generated from some pdf $p(\varepsilon;\eta,\sigma)$ such as the Gaussian distribution with mean 0 and unknown standard deviation σ_i,

$$p(\varepsilon;\eta,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(\varepsilon-\eta)^2}{2\sigma^2}} \ .$$

For example, let $c = 2$ and $l = 2$ then from (1),

$$\begin{aligned} y &= f_1(\mathbf{x};\boldsymbol{\theta}_1) + \varepsilon_1 \\ y &= f_2(\mathbf{x};\boldsymbol{\theta}_2) + \varepsilon_2 \ . \end{aligned}$$

1.1. Mixture Density Estimation

Each $\mathbf{m}_j$ is assumed to be generated with probability $Pr(i)$ from model i such that $\sum_{i=1}^{c} Pr(i) = 1$. The log-likelihood function of the data is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \mathbf{M}) &\doteq \ln \ell(\boldsymbol{\theta}; \mathbf{M}) = \ln Pr(\mathbf{M}|\boldsymbol{\theta}) \\
&= \sum_{j=1}^{d} \ln \sum_{i=1}^{c} Pr(i) \cdot p(\mathbf{m}_j | i, \boldsymbol{\theta}_i) \\
&= \sum_{j=1}^{d} \ln \Big[Pr(i) \cdot p(y_j - \theta_{11} x_j - \theta_{12} | 0, \sigma_1) \\
&\qquad + \big((1 - Pr(i)) \cdot p(y_j - \theta_{21} x_j - \theta_{22} | 0, \sigma_2) \big) \Big]
\end{aligned}
$$

✘ Solution: partition $\mathbf{M} = \{\mathbf{m}_j\}$ into c submodels and find θ_i by regression.

1.2. Fuzzy c-Regression Models

Hathaway, Bezdek [2] :

- ▷ fuzzy partition $\mathbf{M}$.
- ▷ Estimate $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_c$ simultaneously by modifying the fuzzy c-means algorithm (FCM)
- ▷ Assign each $\mathbf{m}_j$ a fuzzy label $u_{ij} \in \mathbf{U}$.
- ▷ Error criterion :

$$E_{ij}[\boldsymbol{\theta}_i] = \|\mathbf{y}_j - f_i(\mathbf{x}_j; \boldsymbol{\theta}_i)\|^2 \ .$$

- ▷ Objective functions:

$$E_w[\mathbf{U}, \{\boldsymbol{\theta}_i\}] = \sum_{i=1}^{c} \sum_{j=1}^{d} u_{ij}^w \cdot E_{ij}[\boldsymbol{\theta}_i] \ ,$$

If the regression functions $f_i(\mathbf{x}; \boldsymbol{\theta}_i)$ are linear in the parameters $\boldsymbol{\theta}_i$, the parameters can be obtained as a solution of the weighted least-squares :

$$\boldsymbol{\theta}_i = \left[\mathbf{X}_e^T \mathbf{W}_i \mathbf{X}_e\right]^{-1} \mathbf{X}_e^T \mathbf{W}_i \mathbf{Y}$$

where (cf. TS parameter identification)

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_d^T \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}, \qquad \mathbf{W}_i = \begin{bmatrix} u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{id} \end{bmatrix}$$

and

$$\mathbf{X}_e = [\mathbf{X}, \mathbf{1}] \,,$$

1.3. Fuzzy c-Regression Algorithm

Preparations:

- ▷ Fix c, $2 \leq c < d$.
- ▷ Choose the termination tolerance $\delta > 0$.
- ▷ Fix w, $1 \leq w < \infty$.
- ▷ Initialise $\mathbf{U}^{(0)} \in M_{fc}$ randomly.

Repeat for $l = 1, 2, \ldots$:

Step 1: Calculate model parameters $\boldsymbol{\theta}_i^{(l)}$ to globally minimise

$$E_w[\mathbf{U}, \{\boldsymbol{\theta}_i\}] = \sum_{i=1}^{c} \sum_{j=1}^{d} u_{ij}^w \cdot E_{ij}[\boldsymbol{\theta}_i] \ ,$$

Step 2: Update $\mathbf{U}$ with $E_{ij} = E_{ij}[\boldsymbol{\theta}_i^{(l-1)}]$, to satisfy:

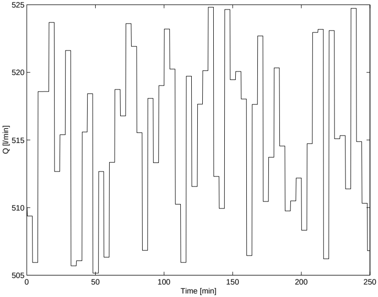
$$u_{ij}^{(l)} = \frac{1}{\sum_{k=1}^{c} \left(\frac{E_{ij}}{E_{kj}} \right)^{\frac{2}{w-1}}}$$

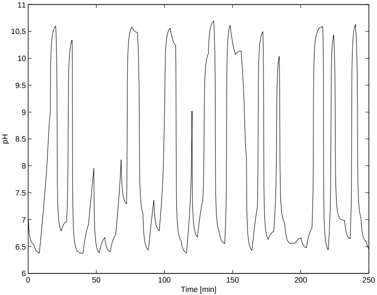
if $E_{ij} > 0$ for $1 \leq i \leq c$, and otherwise $u_{ij} = 0$ if $E_{ij} > 0$, and $u_{ij} \in [0, 1]$ with $(u_{1j} + \cdots + u_{cj}) = 1$.

Until $\left\| \mathbf{U}^{(l)} - \mathbf{U}^{(l-1)} \right\| < \delta$.

2. Example: pH Neutralisation Process

The system consists of a neutralisation tank to which a base Q is added and the pH level is the output variable to be controlled. The identification data set, consisting of $d = 1250$ values:

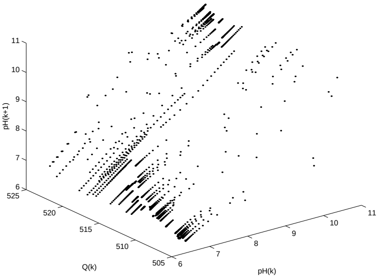


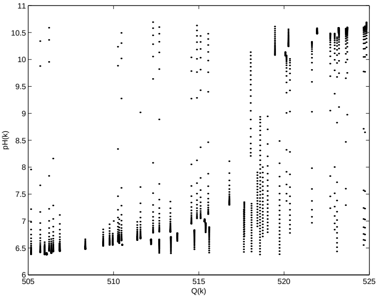


Model structure: First-order discrete time NARX model :

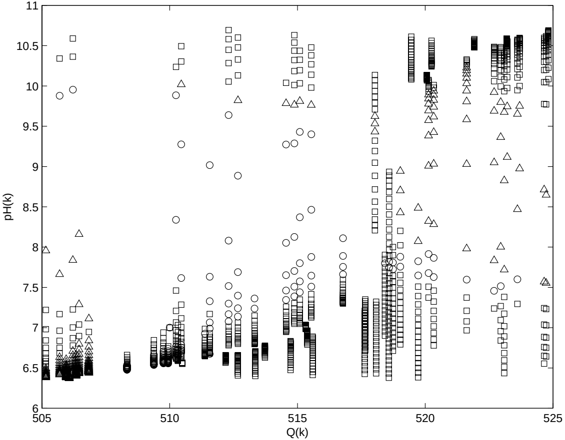
$$pH(k+1) = f\big(pH(k), Q(k)\big)$$

Training data, as used for identification and 90 degrees view:





Fuzzy switching regression model: For $c = 3$, $w = 2$ and the stopping criterion set equal to $\delta = 10^{-10}$. 90 degrees view where points are assigned to the cluster in which they have the maximum membership:



2.1. Discussion

- ✘ Clusters are not immediately apparent.
- ✘ The s-shaped area with a higher density of data corresponds to the equilibrium of the system.
- ✘ The switching regression model does not produce compact sets. Instead it finds the dynamically similar regions on the pH-Q space. This means the points marked by squares are mainly equilibrium points, while the two other clusters cover two different off-equilibrium regions.
- ✘ Hence the method is useful for extracting “fuzzy” knowledge about the possible dynamic regions of the system, but because of the non-compact fuzzy sets one would obtain from the fuzzy clusters, it is difficult transform the result into a rule base.

References

[1] Babuska, R. : *Fuzzy Modelling for Control*. Kluwer, 1998. See http://lcewww.et.tudelft.nl/

[2] Hathaway, R.J. and Bezdek, J.C. : *Switching regression models and fuzzy clustering*. IEEE Transactions on Fuzzy Systems, 3:195–204, August 1993. 5

[3] Wolkenhauer, O. : *Data Engineering*. http://www.csc.umist.ac.uk/people/wolkenhauer.htm.