Non linear robust regression in high dimension
Emeline Perthame, Florence Forbes, Brice Olivier, Antoine Deleforge

To cite this version:
Emeline Perthame, Florence Forbes, Brice Olivier, Antoine Deleforge. Non linear robust regression in high dimension. The XXVIIIth International Biometric Conference, Jul 2016, Victoria, Canada. hal-01423622
Non linear robust regression in high dimension
E. Perthame*, F. Forbes*, B. Olivier*, A. Deleforge**
*MISTIS, INRIA Grenoble, France, **PANAMA, INRIA Rennes, France

1 - Non linear mapping problem
- The goal is to retrieve X from Y through a non linear regression function g
  \[ \mathbb{E}(X|Y = y) = g(y) \]
  with \( Y \in \mathbb{R}^{D}, X \in \mathbb{R}^L, D \gg L \)
  \[ \begin{pmatrix} y_1 \\ \vdots \\ y_D \end{pmatrix} \leftarrow g(Y) \left( \begin{pmatrix} x_1 \\ \vdots \\ x_L \end{pmatrix} \right) = x \]
- For example, Y is a reflectance spectrum \((D = 184)\) measured at a specific location of the Mars planet and X is the composition of the ground at this location \((L = 3)\)

2 - Difficulties
- High dimension \((D \gg L)\) \(\Rightarrow\) Inverse regression strategy
  \[ \mathbb{E}(Y|X = x) = f(x) \]
- Non linear mapping \(\rightarrow\) Piecewise linear approximation of f (and g)
  \[ Y = \sum_{k=1}^{K} \pi_k A_k X + b_k + \varepsilon_k \]
  with \(E(\varepsilon_k^2) \propto \Sigma_k\) and Z multinomial latent variable
  \[ P(Z = k) = \pi_k \]
- Dealing with outliers \(\rightarrow\) Heavy tail distribution
  \(\Rightarrow\) Generalized Student distribution

3 - SLLiM model
- A mixture of Student distributions encodes the piecewise linear regressions
  \[ p(X = x, Y = y|Z = k) = S_{k,D}((x, y)^T; m_k, \varepsilon_k, \alpha_k, 1) \]
  with
  \[ m_k = \begin{bmatrix} c_k \\ A_k c_k + b_k \end{bmatrix} \text{ and } \varepsilon_k = \begin{bmatrix} \Gamma_k \\ A_k \Gamma_k + \Sigma_k + A_k \Gamma_k A_k^T \end{bmatrix} \]
- Therefore, the joint density \((X, Y)\) is a mixture of Student regressions
  \[ p(X = x, Y = y) = \sum_{k=1}^{K} \pi_k S_{k,D}((x, y)^T; m_k, \varepsilon_k, \alpha_k, 1) \]
- We denote by \(\theta = (\alpha_k, \Gamma_k, A_k, b_k, \Sigma_k, \pi_k, \alpha_k)\) \((1 \leq k \leq K)\) the set of parameters
- Extension to partially observed responses

4 - Inverse regression strategy
- Forward strategy \((x = g(y))\), conditionals are
  \[ p(X = x|Y = y; \theta) = \frac{1}{\sum_{k=1}^{K} \pi_k S_{k,D}(y; A_k x + b_k, \Sigma_k, \alpha_k, 1)} \]
  \(\Rightarrow\) D = 500, L = 2, \(\Gamma_k\) diagonal \(\Rightarrow\) 126 254 parameters
- Inverse strategy \((y = f(x))\)
  \[ p(Y = y|X = x; \theta) = \frac{1}{\sum_{k=1}^{K} \pi_k S_{k,D}(x; A_k y + b_k, \Sigma_k, \alpha_k, 1)} \]
  \(\Rightarrow\) D = 500, L = 2, \(\Gamma_k\) diagonal \(\Rightarrow\) 2 003 parameters
- Our approach reduces the number of parameters to estimate
- Prediction : The regression function of interest g is approached by \(\hat{g}\)

5 - Estimation of \(\theta\) by EM algorithm
- E-step
  - E-U step: Update of weight of each data point \(P(U|x, y, Z = k; \theta^{(t)})\)
  - E-Z step: Update posterior probabilities \(P(Z = k|x, y, \theta^{(t)})\)
- M-step
  - \((\pi_k, c_k, \Gamma_k)\) \(\Rightarrow\) Estimation is like a standard Student mixture
  - \((A_k, b_k, \Sigma_k)\) \(\Rightarrow\) Estimation is “Linear regression-like”
  - \(\alpha_k\) \(\rightarrow\) Not in closed-form but standard

6 - Application to air quality in the subway in Paris
- Prediction of NO \((L=1)\) from NO2 \((D=1)\) in Châtelet station in Paris during March 2015 \((N = 341)\) measures
- SLLiM achieves better prediction \(4\) than its Gaussian counterpart (GLLiM) on complete data
- SLLiM is equivalent to GLLiM when no outliers (removed)

7 - Other applications
- Application when \(D \gg L\)
  - Hyperspectral data on Mars
    - D = 184, L = 3, N = 6983
    - K fixed to 10, number of latent variables W estimated by BIC
  - Prediction of proportion of CO2 ice and dust from spectra
    - Near-infrared spectra on orange juices
    - D = 134, L = 1, N = 218
    - Prediction of sucrose level of each orange juice from its spectra
  - Comparison with other non linear regression methods

References
[3] Link to RATP (subway) data: http://data.ratp.fr/explore/dataset/qualite-de-l-air-mesuree-dans-la-station-chatelet

Prop. of CO2 ice Prop. of dust
SLLiM \((K=10)\) 0.168 \((0.019)\) 0.145 \((0.020)\)
GLLiM \((K=10)\) 0.180 \((0.023)\) 0.155 \((0.023)\)
Regression splines 0.173 \((0.016)\) 0.160 \((0.021)\)
SIR 0.243 \((0.025)\) 0.157 \((0.016)\)
RVM 0.299 \((0.021)\) 0.275 \((0.034)\)