CRITICAL EVALUATION OF LLE DATA

Zdeněk Wagner

Orthogonal basis

Vectors in n-dimensional space: $\phi_1, \phi_2, \ldots, \phi_n$

Condition for linear independence:

$$\sum_{i=1}^{n} c_i \phi_i = \emptyset \Longleftrightarrow c_i = 0 \quad \forall i = 1, 2, \ldots, n;\ c_i \in \mathcal{R}^1$$

Orthogonality: $\langle \phi_i | \phi_j \rangle = 0,\ i \neq j; \quad \langle \phi_i | \phi_i \rangle > 0$

Example: $(-2, 0, 0); (0, 2, -1); (0, 3, 6)$

Proof: $\langle (0, 2, -1) | (0, 3, 6) \rangle = 2 \cdot 3 - 1 \cdot 6 = 0$

Vector coordinates

$$F = \sum_{i=1}^{n} c_i \phi_i$$

using condition of orthogonality

$$\langle F | \phi_j \rangle = \sum_{i=1}^{n} c_i \langle \phi_i | \phi_j \rangle = c_j \langle \phi_j | \phi_j \rangle$$

$$c_j = \frac{\langle F | \phi_j \rangle}{\langle \phi_j | \phi_j \rangle}$$

Vector space of continuous functions

Functions $\phi_0(x), \phi_1(x), \phi_2(x), \ldots$ **continuous on** $x \in [a, b]$, $a, b \in \mathcal{R}^1$, $\phi_i(x) \in \mathcal{R}^1$.

$$F = \sum_{i=0}^{\infty} c_i \phi_i \Longleftrightarrow F(x) = \sum_{i=0}^{\infty} c_i \phi_i(x), \quad x \in [a, b]$$

$$\langle \phi_i | \phi_j \rangle = \int_a^b w(x) \phi_i(x) \phi_j(x) dx, \quad w(x) \geq 0$$

$$\langle F | \phi_j \rangle = \sum_{i=0}^{\infty} c_i \langle \phi_i | \phi_j \rangle = c_j \langle \phi_j | \phi_j \rangle$$

$$\langle \phi_i | \phi_j \rangle' = \sum_{k=1}^{K} w(x_k) \phi_i(x_k) \phi_j(x_k), \quad w(x_k) \geq 0$$

Legendre polynomials

Polynomials $P_n(x)$ orthogonal on $x \in [-1, 1]$ with $w(x) = 1$.

$$P_0(x) = 1$$

$$P_1(x) = x$$

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

$$\langle P_m | P_n \rangle = \int_{-1}^{1} P_m(x)P_n(x)dx = \frac{2}{2n+1}\,\delta_{mn}$$

https://en.wikipedia.org/wiki/Legendre_polynomials

Clenshaw Recurrence Formula

$$F(x) = \sum_{i=0}^{n} c_i \phi_i(x), \quad x \in [a, b]$$

- Linear combination $F(x)$ of $\phi_i(x)$ cannot be precisely evaluated due to rounding errors.
- Evaluation carried out by a recurrence formula using the same coefficients as in the recurrent definition of $\phi_i(x)$.

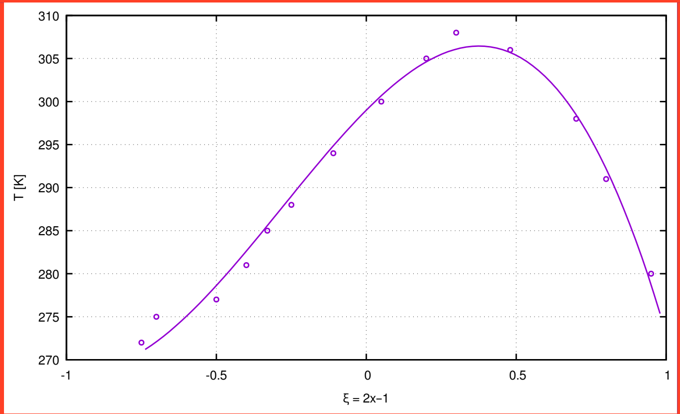
https://mathworld.wolfram.com/ClenshawRecurrenceFormula.html

Application to regression

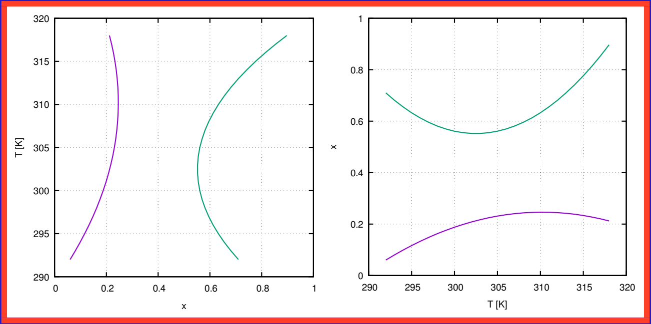
$$\xi_k = \frac{2x - x_{\mathsf{min}} - x_{\mathsf{max}}}{x_{\mathsf{max}} - x_{\mathsf{min}}}$$

- Function `GrregX` requires matrix `x` and vector `y` where `y` is a vector of experimental values of the dependent variable and `x` is a matrix of Legendre polynomials up to the selected degree evaluated at transformed values ξ_k.
- Matrix `x` is calculated by function `mklegendrematrix`.
- Clenshaw formula for Legendre polynomials is available in function `legendreclenshaw`.
- A value of a Legendre polynomial of given order is calculated by function `legendrepol` using the recurrence formula.

Cloud points without known LCST/UCST



Generic case



$$x_A = f_A(\tau), \quad x_B = f_B(\tau), \quad \tau = \frac{2T - T_{\text{min}} - T_{\text{max}}}{T_{\text{max}} - T_{\text{min}}}$$

Method

1. Choose the type of regression, either $x = f(\tau)$ or $T = f(\xi)$
2. Perform regression for each set separately and for all sets
3. Marginal analysis of residuals for each set and for everything: EGDF, additive, heteroscedastic
4. Identify outliers for each set and for all sets
5. If there are too many outliers for all combined data, repeat regression and marginal analysis of inliers and outliers in order to obtain homogeneous subsamples with just a few (or no) outliers

… continued

6. Compare the sizes of tolerance intervals and intervals of typical data
7. Calculate residuals of all data for results of regresion of each set and all data
8. Plot residuals as a function of T or x (depending on the type of regression) and mark the gnostic intervals including LB and UB (if they are not too far)