



ROZSZERZANIE BAZY I REGULARYZACJA

KATARZYNA FRANKIEWICZ



WPROWADZENIE DO FUNKCJI BAZOWYCH

$$f(X) = \sum_{m=1}^M \beta_m h_m(X)$$

Gdzie funkcje h to funkcje bazowe, np.:

$$h_m(X) = X_m, \text{ gdzie } m=1, \dots, p$$

$$h_m(X) = X_j^2$$

$$h_m(X) = X_j X_k$$

$$h_m(X) = \log(X_j)$$

$$h_m(X) = \sqrt{X_j}$$

$$h_m(X) = \|X\|$$

$$h_m(X) = \mathbb{1}(L_m \leq X_k < U_m)$$

SPLINE

- Splines to funkcje, które składają się z prostszych funkcji, które nazywamy funkcjami bazowymi.
- Kiedy używamy spline'ów każda z funkcji bazowych b_k ma pewien współczynnik β_k , czyli

$$s(x) = \sum \beta_k b_k(x)$$

POLYNOMIAL AND SPLINES

- Dla ułatwienia zrozumienia tematu, w dalszej części przyjmujemy, że wektor X jest wektorem jednowymiarowym.

$$h_1(X) = I(X < \xi_1), \quad h_2(X) = I(\xi_1 \leq X < \xi_2), \quad h_3(X) = I(\xi_2 \leq X)$$

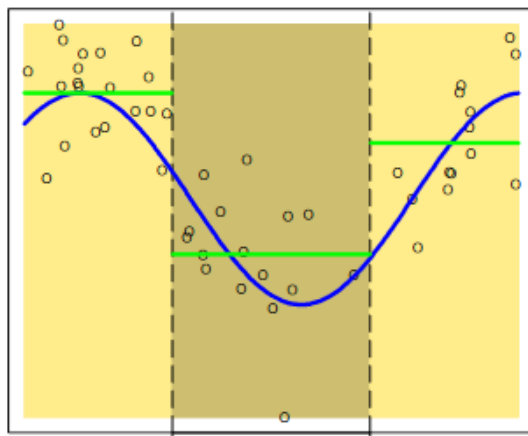
■

$$\beta_1 + \xi_1\beta_4 = \beta_2 + \xi_1\beta_5$$

$$\beta_2 + \xi_2\beta_5 = \beta_3 + \xi_1\beta_6$$

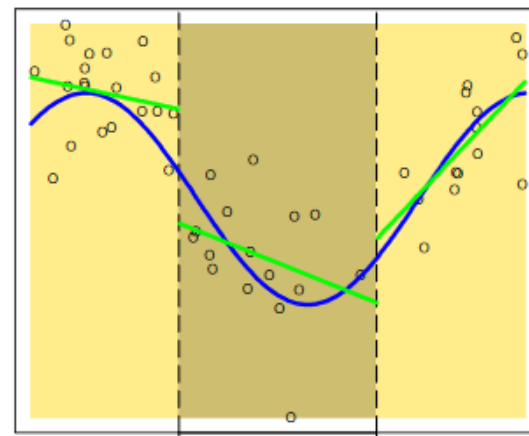
$$h_1(X) = 1, \quad h_2(X) = X, \quad h_3(X) = (X - \xi_1)_+, \quad h_4(X) = (X - \xi_2)_+$$

Piecewise Constant



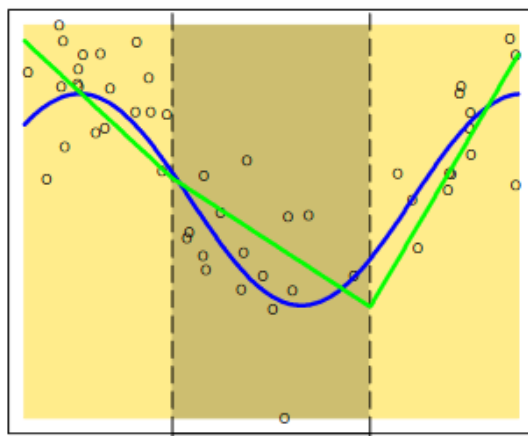
ξ_1 ξ_2

Piecewise Linear



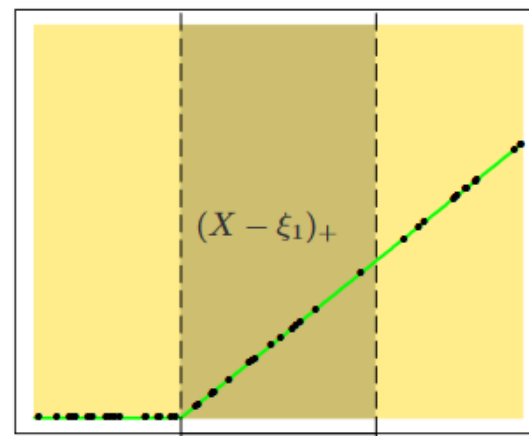
ξ_1 ξ_2

Continuous Piecewise Linear

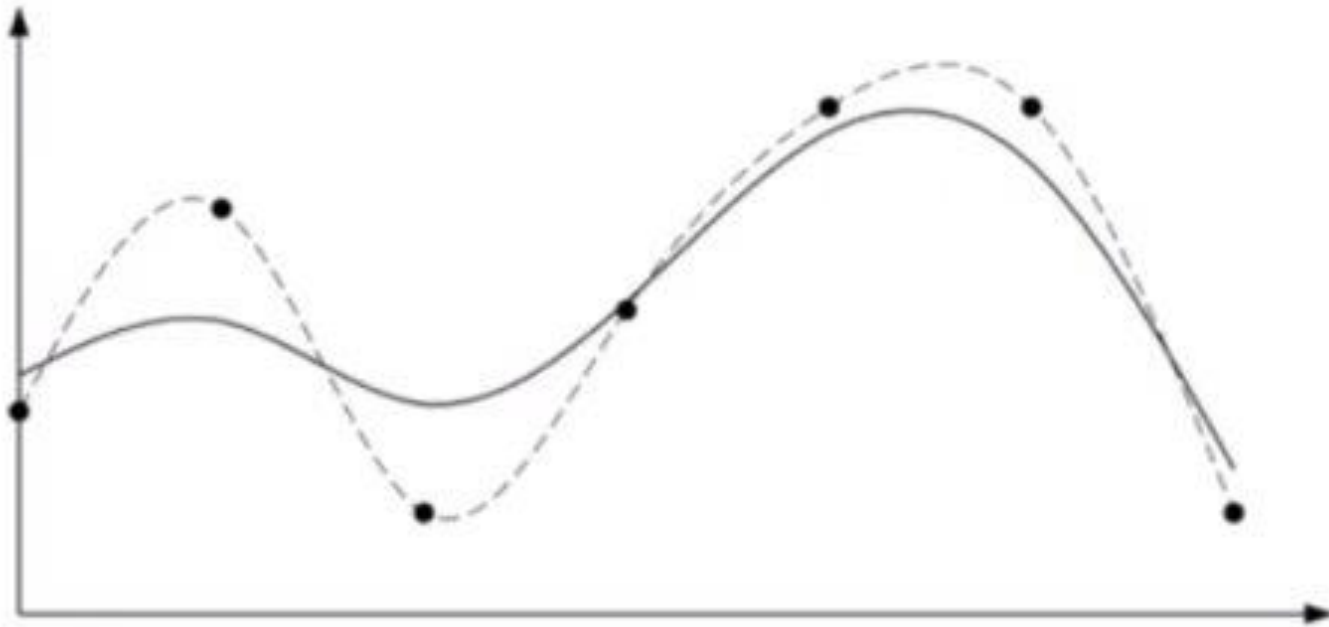


ξ_1 ξ_2

Piecewise-linear Basis Function



ξ_1 ξ_2



CUBIC
SMOOTHING
SPLINE
I NATURAL
CUBIC SPLINES

$$\begin{aligned}
 h_1(X) &= 1, & h_3(X) &= X^2, & h_5(X) &= (X - \xi_1)_+^3, \\
 h_2(X) &= X, & h_4(X) &= X^3, & h_6(X) &= (X - \xi_2)_+^3.
 \end{aligned}$$

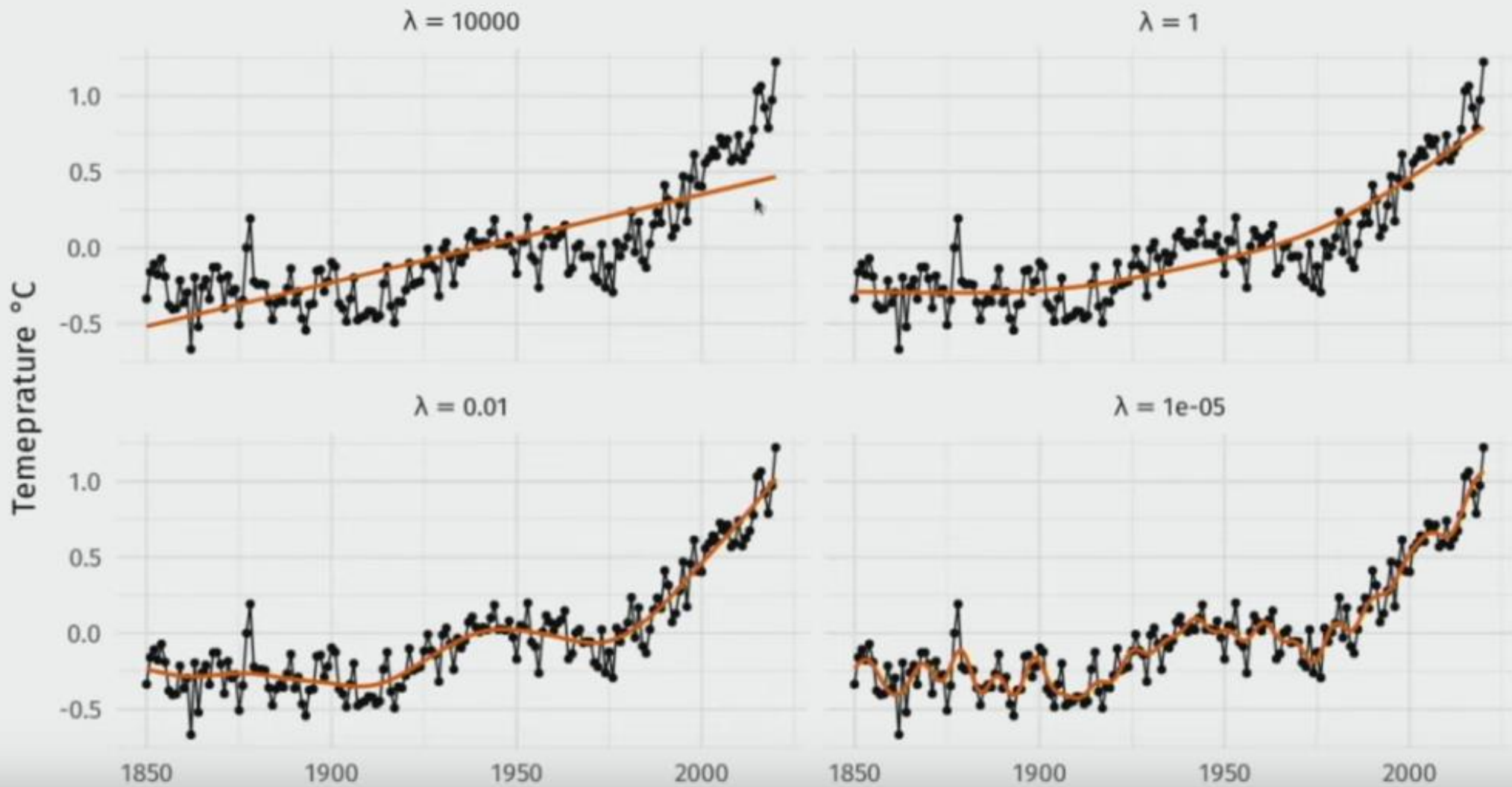
SMOOTHING SPLINE

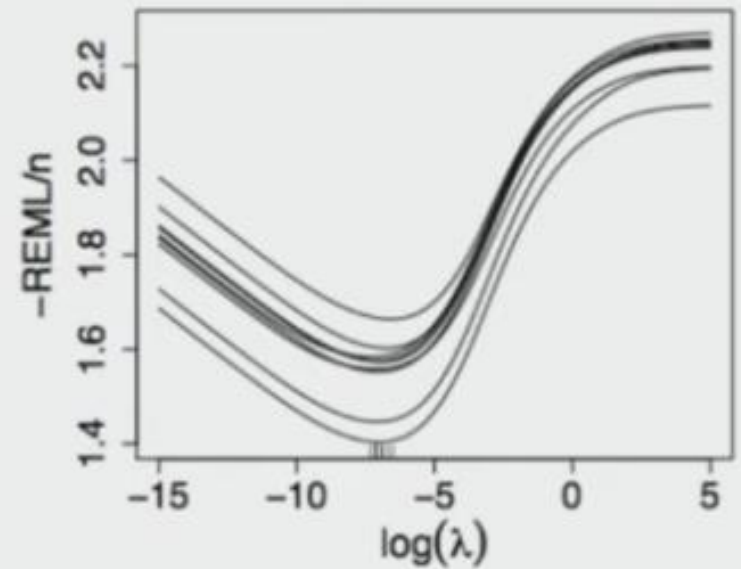
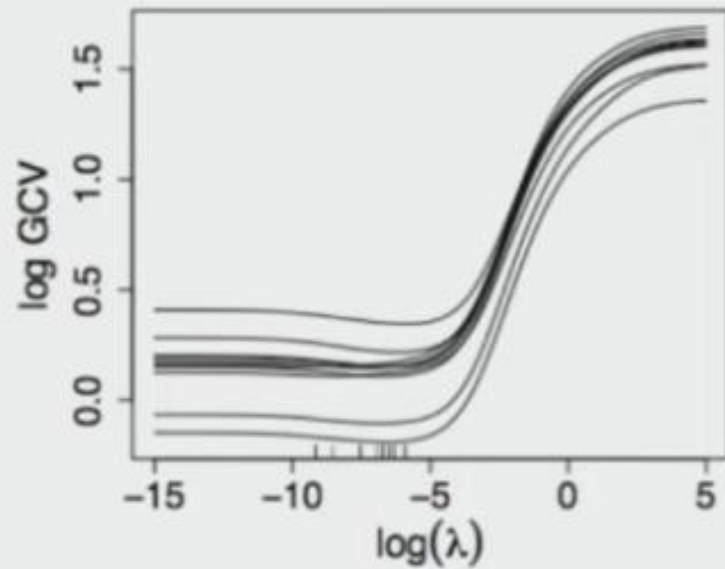
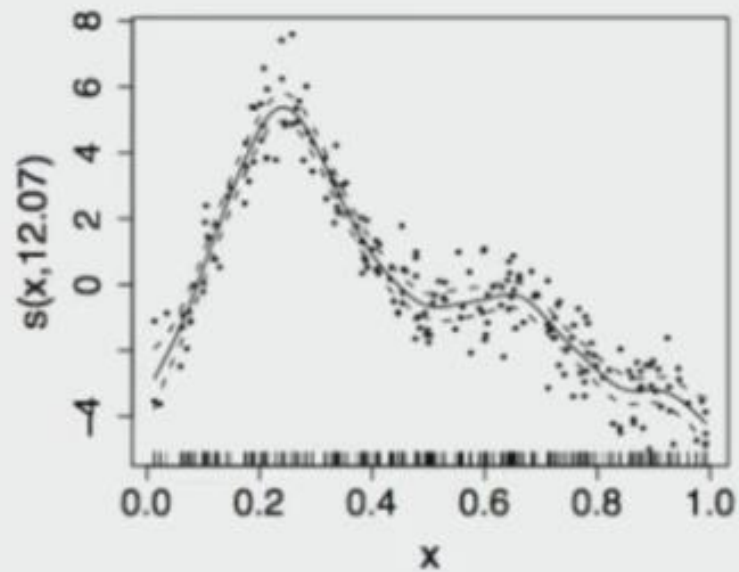
$$\text{RSS}(f, \lambda) = \sum_{i=1}^N \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt$$

WYBÓR PARAMETRU WYGŁADZENIA

$$\int_{\mathbb{R}} [f'']^2 dx = \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta} = W$$

$$\mathcal{L}_p = \log(\text{Likelihood}) - \lambda W$$





WYBÓR PARAMETRU WYGŁADZENIA

MAXIMUM ALLOWED WIGGLINESS

- Wartość parametru k musi być wystarczająco duża, a parametr wygładzenia zajmie się resztą dopasowania modelu do danych.
- Wystarczająco duża oznacza, że wartość parametru k powinna pokryć domyślną złożoność modelu.

DZIĘKUJĘ ZA UWAGĘ!

