Overview

Motivation: data points might be affected by errors. The interpolation criterion $y_i = f(x_i) = g(x_i)$, i.e., exactly recover the function values is not appropriate, and linear combination of basis functions is not appropriate

- Approximants with nonlinear dependence on parameters
- Gradient descent

- Problem: approximate $f\colon [a,b] \to \mathbb{R}$ by $g\colon [a,b] \to \mathbb{R}$, $f(t) \cong g(t;c)$, c are approximant parameters
- Type of approximant dependence on parameters.
  - linear: $g = c_1 g_2 + \cdots + c_n g_n$, typically arising from unknown behavior.
  - nonlinear: $g(t;c)$, typically arising from known behavior. Examples:
    - $\longrightarrow$ Chemical reaction rate given by Arrhenius equation (linearizable)

$$k(T) = A \exp\left[-\frac{E}{RT}\right]$$

Data $\mathcal{D} = \{(T_i, k_i), i = 1, \ldots, m\}$, A, E are parameters, R is a constant.

- $\longrightarrow$ Statistical model parameters, e.g., bimodal model

$$g(x;c) = A_1 \exp\left[-\frac{(x-\mu_1)}{\sigma_1^2}\right] + A_2 \exp\left[-\frac{(x-\mu_2)}{\sigma_2^2}\right]$$

$\mathcal{D} = \{(x_i, g_i), i = 1, \ldots, m\}$, parameters: $\boldsymbol{c} = [\ A_1 \ \ A_2 \ \ \mu_1 \ \ \mu_2 \ \ \sigma_1 \ \ \sigma_2\ ]^T$

- Chemical reaction rate given by Arrhenius equation (linearizable)

$$k(T) = A\exp\left[-\frac{E}{RT}\right]$$

Data $\mathcal{D} = \{(T_i, k_i), i = 1, \ldots, m\}$, A, E are parameters, R is a constant.

- Take logarithm

$$y = \log k = c_1 - c_2 x, x = \frac{1}{T}, c_1 = \log A, c_2 = E/R$$

$$\mathcal{L} = \{(x_i, y_i), i = 1, \ldots, m\}$$

- Obtain linear dependence on parameters, apply projection technique (Lesson11)

- Bimodal distribution

$$g(x;c) = A_1 \exp\left[-\frac{(x-\mu_1)}{\sigma_1^2}\right] + A_2 \exp\left[-\frac{(x-\mu_2)}{\sigma_2^2}\right]$$

$\mathcal{D} = \{(x_i, y_i = g(x_i; c)), i = 1, \ldots, m\}$, parameters: $\boldsymbol{c} = [\ A_1\ \ A_2\ \ \mu_1\ \ \mu_2\ \ \sigma_1\ \ \sigma_2\ ]^T$

- Taking logarithms does not lead to a linear system
- Form squared error in 2-norm

$$f(\boldsymbol{c}) = \varepsilon^2 = \| g(\boldsymbol{x}; \boldsymbol{c}) - \boldsymbol{y} \|_2^2$$

- Solve minimization problem

$$\min_{\boldsymbol{c} \in \mathbb{R}^n} f(\boldsymbol{c})$$

- To solve $\min_{\boldsymbol{c} \in \mathbb{R}^n} f(\boldsymbol{c})$, $f : \mathbb{R}^n \to \mathbb{R}$, $f > 0$, f convex recall that the gradient defines direction of maximal inrease of f
- Gradient descent. Construct a sequence of parameter approximants $\{\boldsymbol{c}_k\}_{k \in \mathbb{N}}$
- Starting from some previous approximation $\boldsymbol{c}_k$, new estimate of the parameters obtained by

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \lambda_k \operatorname{grad} f(\boldsymbol{c}_k)$$

- Distance to travel in direction of gradient typically determined by orthogonality to next gradient

$$\operatorname{grad} f(\boldsymbol{c}_k) \cdot \operatorname{grad}[f(\boldsymbol{c}_{k+1})] = 0 \Rightarrow \text{equation forspace} \lambda_k$$

- Topic pursued further during solution of root finding.