Generalized additive model (GAM)

1 Simple to complex models

1.1 Linear Models (LM)

$$y \sim \mathcal{N}(\mu, \sigma^2) \tag{1}$$

$$E(y) = \mu = X\beta \tag{2}$$

where y is the response vector, X is the design matrix, β is the vector of regression coefficients

1.2 Generalized Linear Models (GLM)

1. A probability distribution from the exponential family $y \sim ExpoFamily(\mu, etc)$
2. A linear predictor $\eta = X\beta$
3. A link function g such that $E(Y) = \mu = g^{-1}(\eta)$

For more details, see `https://en.wikipedia.org/wiki/Generalized_linear_model`

For example

y is a continuous variable which follows a Gaussian distribution and g is an identity link function, then the GLM returns to LM.

$$y \sim \mathcal{N}(\mu, \sigma^2) \tag{3}$$

$$E(y) = \mu = g^{-1}(\eta) = X\beta \tag{4}$$

y is a count variable which follows a Poisson distribution and g is a log link function.

$$y \sim \mathcal{P}(\mu) \tag{5}$$

$$E(y) = \mu = g^{-1}(\eta) = e^{X\beta} \tag{6}$$

y is a categorical variable which follows a Categorical distribution and g is a logit link function

$$y \sim Cat(\mu) \tag{7}$$

$$E(y) = \mu = g^{-1}(\eta) = \frac{e^{X\beta}}{1+e^{X\beta}} \tag{8}$$

1.3 Generalized Addictive Model (GAM)

1. A probability distribution from the exponential family $y \sim ExpoFamily(\mu, etc)$
2. A linear predictor $\eta = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m)$
3. A link function g such that $E(Y) = \mu = g^{-1}(\eta)$

where f_i may be smooth functions with a specified parametric form (polynomial, un-penalized regression spline of a variable) or may be specified non-parametric form.

Usually, people use R MGCV package to run GAM, for details please see `https://m-clark.github.io/docs/GAM.html`. One important thing is to choose the smooth function f, the default function of MGCV package is low rank thin plate spline function.
Consider the problem of estimating the smooth function $f(x)$ where x is a d-vector, from n obervations (y_i, x_i) such that

$$y_i = f(x_i) + \epsilon_i \tag{9}$$

where ϵ_i is a random error term. Thin plate splines can be used to estimate f by finding the function g minimizing

$$\|y - g\|^2 + \lambda J_{md}(g) \tag{10}$$

where y is the vector of y_i data, $g = (g(x_i), g(x_2), \cdots, g(x_n))^{'}$, $J_{md}(g)$ is a penalty functional measuring the wiggliness of g and λ controls the trade-off between data fitting and smoothness of g. The wiggliness penalty is defined as

$$J_{md} = \int \cdots \int_{R^d} \sum_{v_1+\cdots+v_d=m} \frac{m!}{v_1! \cdots v_d!} (\frac{\partial^m g}{\partial x_1^{v_1} \cdots x_d^{v_d}})^2 dx_1 \cdots dx_d \tag{11}$$

Provided that we impose the technical restriction $2m > d$, it can be shown that the function minimizing above equation (10) has the form

$$g(x) = \sum_{i=1}^{n} \delta_i \eta_{md}(\|x - x_i\|) + \sum_{j=1}^{M} \alpha_j \phi_j(x) \tag{12}$$

where δ and α are unknown parameter vectors subject to the constraint that $T^{'}\delta = 0$ and $T_{ij} = \phi_j(x_i)$. The $M = \binom{m+d-1}{d}$ function ϕ_i are linearly independent polynomials spanning the space of polynomials in R^d of degree less than m.

$$\eta_{md}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d} \log(r), & d \text{ even} \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d}, & d \text{ odd} \end{cases} \tag{13}$$

For more details of the smoothing function, see paper ”Thin plate regression splines”.