Generalized Linear Models

Generalized Linear Models Can Be Used As Surrogate Models

Definition: Generalized Linear Models

A generalized linear model (GLM) is a surrogate model of the form

$$\hat{f}(x, \alpha) = \sum_{k=1}^{p} \alpha_k \phi_k(x)$$

where $\{\phi_k(x)\}_{k=1}^{p}$ is a fixed set of basis functions

- GLM is linear in the parameters α_k
- The basis functions do <u>not</u> need to be linear.

Perhaps the Simplest GLM Is One With a Linear Basis

Definition: GLM with a linear basis

A generalized linear model with a linear basis takes the form

$$\hat{f}(x, \alpha) = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k,$$

$p = n + 1$

which defines a hyperplane in $\mathbb{R}^n$.

- A GLM with a linear basis has $(n+1)$ parameters.

Parameter Estimation: How Do We Determine the α_i for GLMs?

Suppose the data generation step (e.g. LHS) has produced s samples

$$\{(x^{(j)}, f^{(j)})\}_{j=1}^{s}.$$

$f^{(j)} = f(x^{(j)})$

Definition: Interpolating Model

We say the surrogate model $\hat{f}$ interpolates the data if

$$\hat{f}(x^{(j)}, \alpha) = f(x^{(j)}), \qquad \forall j = 1, 2, \ldots, s.$$

The Interpolation Condition Can Be Written Succinctly In Matrix Notation

for a linear basis in a GLM

$$V\alpha = y$$

where

each basis function gets a column

each sample gets a row

$$V = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(s)} & x_2^{(s)} & \cdots & x_n^{(s)} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} f(x^{(1)}) \\ f(x^{(2)}) \\ \vdots \\ f(x^{(s)}) \end{bmatrix}$$

- V is called the Vandermonde matrix

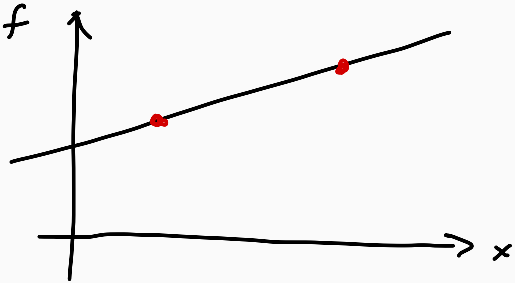
$$\hat{f}(x^{(j)};\alpha) = \underset{\uparrow (1)}{\alpha_0} + \alpha_1 x_1^{(j)} + \alpha_2 x_2^{(j)} + \cdots + \alpha_n x_n^{(j)}$$

The Parameters α_i Can Often Be Determined Using Interpolation

Assuming $s = n + 1$, and the sample locations $\{x^{(j)}\}_{j=1}^{s}$ are unique, then

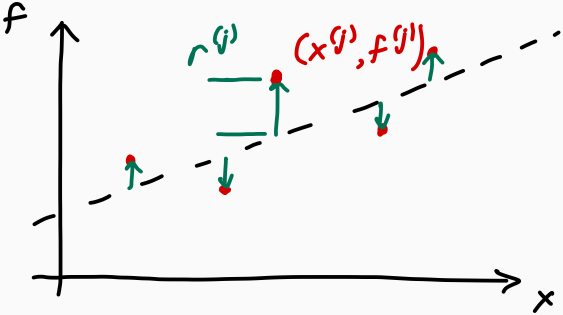
$$\alpha = V^{-1} y$$

- in this case we interpolate the data $\hat{f}(x^{(j)}, \alpha) = f(x^{(j)})$



What If We Have More Data Points Than Parameters?

- For GLMs with linear basis functions, if $s > n + 1$, then it is impossible to interpolate all of the points.
- Instead of interpolating the data, we can seek a least-squares fit of the data.



A Least-Squares Fit Involves Minimizing the Residual Vector

For each data point, $x^{(j)}$, we define the residual

$$r^{(j)}(\alpha) = \hat{f}(x^{(j)}, \alpha) - f(x^{(j)})$$

← general form of residal

$$= \left(\alpha_0 + \sum_{k=1}^{n} \alpha_k x_k^{(j)}\right) - f(x^{(j)}) \neq 0$$

linear basis

Gathering all of the residuals, we can define the residual vector as

$$R(\alpha) = V\alpha - y \neq 0.$$

The Least-Squares Optimization Can Be Solved With Linear Algebra

$$\min_{\alpha} f(\alpha) = \frac{1}{2} R(\alpha)^T R(\alpha). \quad = \frac{1}{2}(V\alpha - y)^T(V\alpha - y)$$

$$\frac{\partial f}{\partial \alpha_j} = \frac{\partial}{\partial \alpha_j} \frac{1}{2} \sum_{i=1}^{s} R_i(\alpha) R_i(\alpha) = \sum_{i=1}^{s} R_i(\alpha) \frac{\partial R_i}{\partial \alpha_j} \quad \frac{\partial R_i}{\partial \alpha_j} = V_{ij}$$

$$\frac{\partial f}{\partial \alpha} = R(\alpha)^T V = (V\alpha - y)^T V = 0 \implies V^T V \alpha = V^T y$$

$$\nabla_\alpha^2 f = V^T V \Leftarrow \text{Hessian is positive def}$$

$$\therefore \quad \alpha = (V^T V)^{-1} V^T y$$

is unique.