Let $f(x) = E(Y \mid X = x)$,

our true nonlinear regression fun;

x : vector of covariates for Y. $\dim(x) = p$

Let's do Taylor expansion to approx. f:

let x_0 be some "central/typical" value of x (e.g., sample mean of $\underbrace{x_1, \dots, x_n}_{\dim(x_i) = p}$ for training data).

gradient/vector of partial deriv's wrt x

$$f(x) \approx f(x_0) + \left[\nabla f(x_0)\right]^T (x - x_0)$$
$$+ \frac{1}{2} \cdot (x - x_0)^T H(x_0)(x - x_0)$$
$$+ \cdots$$

$H(x_0)$: matrix of second derivatives of $\underbrace{f}_{\text{wrt } x}$ at point x_0 (Hessian).

In linear approx., $\nabla f(x_0) \equiv \beta$ - vector of regression coefficients in linear models.

In quadratic approximation, terms of the form $x_i \cdot x_j$ correspond to interactions.

jth component of x.

ith component of x.