Duan, Naihua; Li, Ker-Chau

A bias bound for least squares linear regression. (English) Zbl 0824.62057

Stat. Sin. 1, No. 1, 127-136 (1991).

Summary: Consider a general linear model $y = g(\alpha + \beta\mathbf{x}) + \varepsilon$, where the link function g is arbitrary and unknown. The maximal component of (α, β) that can be identified is the direction of β, which measures the substitutibility of the components of $\mathbf{x}$. If $\zeta(\beta\mathbf{x}) = E(\mathbf{x} \mid \beta\mathbf{x})$ is linear in $\beta\mathbf{x}$, the least squares linear regression of y on $\mathbf{x}$ gives a consistent estimate for the direction of β, despite possible nonlinearity in the link function. If $\zeta(\beta\mathbf{x})$ is nonlinear, the linear regression might be inconsistent for the direction of β.

We establish a bound for the asymptotic bias, which is determined from the nonlinearity in $\zeta(\beta\mathbf{x})$, and the multiple correlation coefficient R^2 for the least squares linear regression of y on $\mathbf{x}$. According to the bias bound, the linear regression is nearly consistent for the direction of β, despite possible nonlinearity in the link function, provided that the nonlinearity in $\zeta(\beta\mathbf{x})$ is small compared to R^2. Our measure of nonlinearity in $\zeta(\beta\mathbf{x})$ is analogous to the maximal curvature studied by *D. R. Cox* and *N. J. H. Small* [Biometrika 65, 263-272 (1978; Zbl 0386.62041)]. The bias bound is tight; we give the construction for the least favorable models which achieve the bias bound. The theory is applied to a special case for an illustration.

MSC:

62J05 Linear regression; mixed models
62J12 Generalized linear models (logistic models)

Keywords:

lack of fit; projection index; projection pursuit; collinearity; general linear model; link function; least squares; asymptotic bias; multiple correlation coefficient; bias bound; nonlinearity; maximal curvature; least favorable models