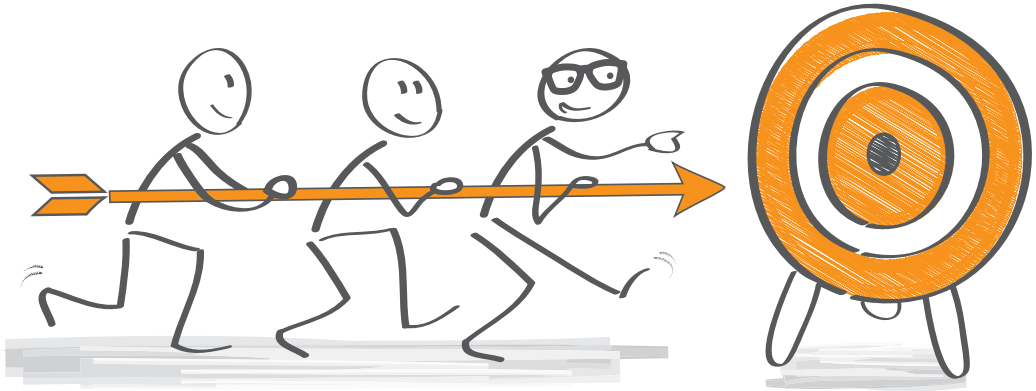
Regression Analysis
Other Regression Methods

Nicoleta Serban, Ph.D.
Associate Professor
Stewart School of Industrial and Systems Engineering

Nonlinear & Nonparametric Regression

GTx

About this lesson



Multiple Linear Regression

What if nonlinearity assumption does not hold?

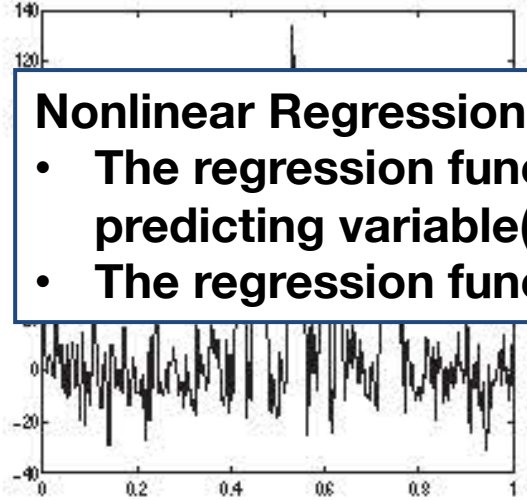
- **If it does not hold for few quantitative predicting variables and transformation not known ⇒ Use transformations of the predicting variables to improve fit**
- **If the relationship between response and the predicting variables is known ⇒ Use Nonlinear Regression**
- **If it does not hold for many variables ⇒ Use Generalized Additive Regression**

- *Independence Assumption*: $\{\varepsilon_1, ..., \varepsilon_n\}$ are independent random variables
- *Normality Assumption*: $\varepsilon_i \sim$ Normal

Example: Nonlinear Regression

- Mass spectroscopy or NMR frequency data:

$$Y_i = \sum_{l=1}^{L} \frac{A_j}{\frac{(x_i - \mu_l)^2}{\tau^2} + 1} + \varepsilon_i$$



- Each peak corresponds to one component

Nonlinear Regression

- **The regression function has a known structure given the predicting variable(s); and**
- **The regression function depends on a series of parameters.**

the peaks

  - A_l for l=1,..,L which are the amplitudes of the peaks

Nonlinear Regression

Nonlinear vs Linear Regression

- **Use least squares approach for estimation;**
- **Assume same assumptions on the error terms hence goodness of fit can be performed similarly;**
- **Regression function is nonlinear vs linear in the parameters;**
- **Estimation of the parameters not in close form expression for nonlinear regression.**
- **R software: nls() vs lm() ⇐ Nonlinear regression is more challenging to implement**

$$\rangle \quad (Y_i - \mathrm{f}(\mathrm{xi}_{1,\ldots,}\mathrm{xip}|\,\theta))^2$$

- Apply numeric algorithms to obtain the estimate for θ

Nonparametric Regression

Model Description:

Nonparametric Regression

- **The regression function has an unknown structure given the predicting variable(s); and**
- **The regression function does not depend on parameters.**

Model Estimation:

- *Curse of dimensionality*: To maintain a given degree of accuracy of an estimator, the sample size must increase exponentially with the dimension p.
    - n = 30000 points when d = 5 to get the same accuracy as n = 300 when d = 1.

Nonparametric Regression (cont'd)

Nonparametric vs Linear Regression

- **The relationship of a predicting variable to the response is assumed unknown**
- **Estimation using the least squares**
- **Estimation of the parameters not in close form expression**
- **R software: gam() in mgcv or gam library**

Model Estimation:

- Backfitting algorithm:
  - Initialize: $\hat{\alpha}$, $\hat{f}_1$,..., $\hat{f}_p$
  - Iterate until convergence: For j=1,..,p

$\check{R}_i = \mathrm{Y_i} - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(x_{ki})$ and estimate $\hat{f}_j$ from regressing $\check{R}_i \sim x_{ji}$

Summary