8 - Generalized Linear Models

Note Title 10/11/2011

- The CLM
  - meaning
  - how fit
  - failings
- The GLM (GLZ)
  - meaning
  - how fit
  - needs
  - failings

Model selection
- R^2
- adjusted R^2
- AIC
- others

The CLM

Classical Linear Model

Dependent Variable

Independent Variables

Object: Summarize or explain depvar with a line.

Method: OLS = Ordinary least squares

Assumptions of CLM and OLS

- effects are linear
- residuals are Normally distributed

implies

$$= \quad y = \hat{\beta}_0 + \hat{\beta} x + e$$

$$e \sim N(0, \sigma^2)$$

equivalent

$$\Rightarrow \quad y \sim N(\beta_0 + \beta x, \sigma^2)$$

Generalized Linear Models GLMs

1. Linear predictor

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$$

Table 6.1
p 181

2. Distribution of y-variable
   - today: Gaussian
     - continuous
     - unbounded
     - constant variance

3 The link function

Note: $\eta = \beta_0 + \beta_1 x_1 + \dots$

Also: μ

link function: $g(\mu) = \eta$

for Gaussian, g is usually the identity

as $\mu = \eta$

Model selection

Adjusted R^2 bigger is better

- penalizes you for adding an additional x-variable.

AIC - smaller is better

- penalizes for adding an additional x-variable
- guideline as no known distributin

Other selection criteria:

SBC = • BIC - Bayesian Information Criteria

Smaller is better

- penalizes for additional x-vars

- penalizes more than AIC

$$AIC = L + 2p$$

$$BIC = L + p \ln n$$

n: data size

Today's thrust:

3 requirements to use GLMs:

1. Linear predictor
   $\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

2. Distribution of dependent variable
   - specifically: its expected value μ

3. Link $g(\mu) = \eta$