Generalized Linear Models – Part I

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 11 – Lecture 11B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Generalized Linear Model

Generalized linear model (GLM) is a unification of linear and nonlinear regression models that allows incorporation of non-normal response distribution.

In a GLM, the response variable distribution must be a member of exponential family, such as normal, Poisson, binomial, exponential, gamma, etc.

Logistic Regression

Model with a binary response

$y_i = 1$ (success) or 0 (failure), $i = 1, \ldots., n$

$P(y_i = 1) = \pi_i$

$P(y_i = 0) = 1 - \pi_i$

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

Logistic Regression

Now suppose that the "success" probability π_i depends on a set of regressors $\mathbf{x}_i$ and the model is

$$E(y_i) = \mathbf{x}_i'\boldsymbol{\beta} = \pi_i$$

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

$$\varepsilon_i = 1 - \mathbf{x}_i'\boldsymbol{\beta} \quad \text{when } y_i = 1$$

$$\varepsilon_i = -\mathbf{x}_i'\boldsymbol{\beta} \quad \text{when } y_i = 0$$

Logistic Regression

Some basic problems with this model

1) ε takes only two values, cannot be normal

2) Error variance is not constant

$$\sigma_{y_i}^2 = E\{y_i - E(y_i)\}^2 = (1-\pi_i)^2 \pi_i + (0-\pi_i)^2 (1-\pi_i) = \pi_i (1-\pi_i)$$

$$\sigma_{y_i}^2 = E(y_i)[1 - E(y_i)]$$

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} = \pi_i$$

Error function is a function of the mean and regressors

Logistic Regression

3) There is a constraint on the response function, because,

$$0 \leq E(y_i) = \pi_i \leq 1$$

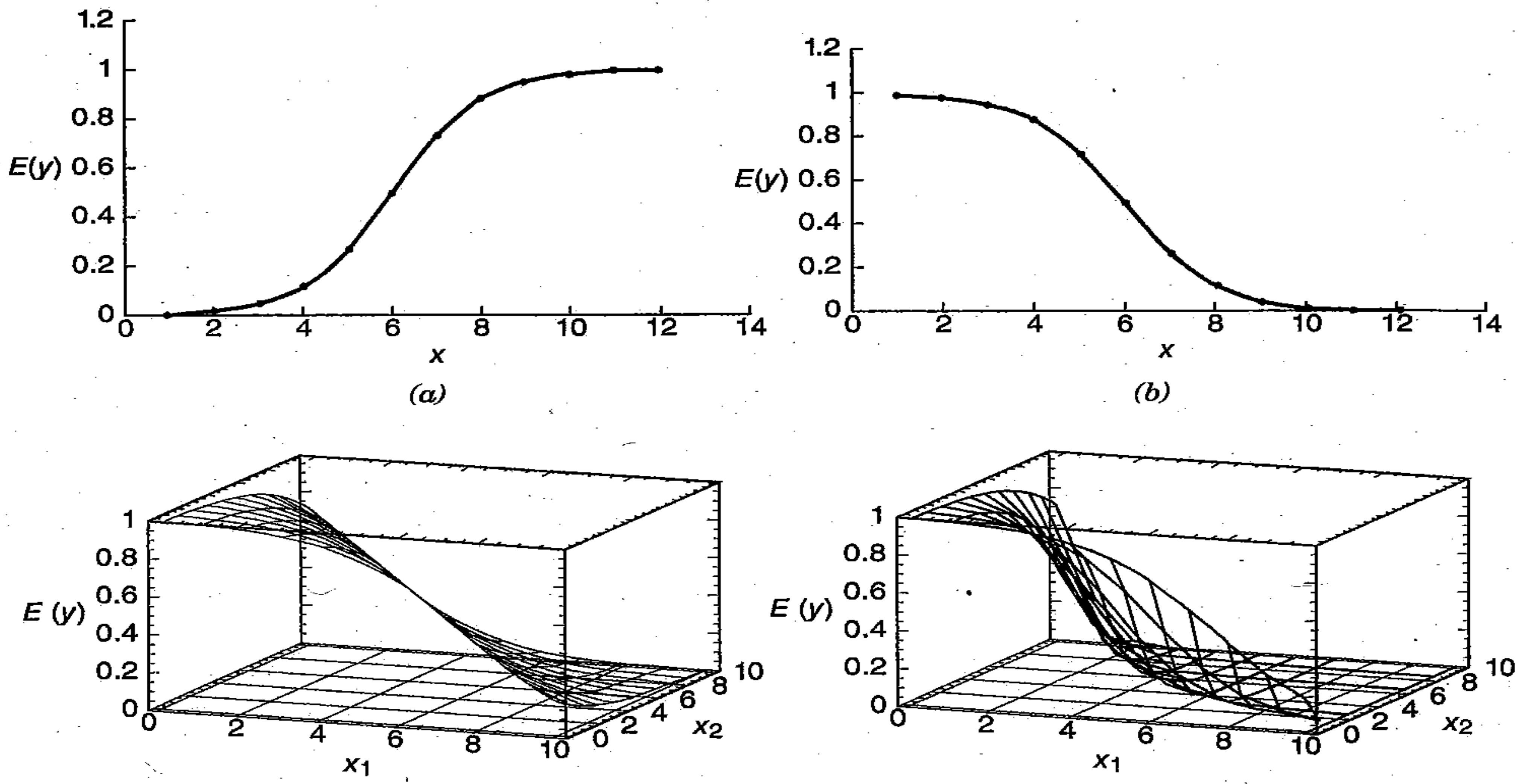
This constraint can cause a serious problem with the choice of linear response function. The fitted linear model may generate predicted responses to lie outside of 0 or 1.

Logistic Regression

Generally, when the response is binary, the shape of the response function should be nonlinear. A monotonically increasing (or decreasing) S-shape response function is usually employed. This response function is called "logistic response function":

$$E(y) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})}$$

Logistic Regression



Logistic Regression

The logistic response function can be linearized.

$$\eta = \mathbf{x}'\boldsymbol{\beta} \quad \Leftarrow \text{linear predictor}$$

$$\eta = \ln\frac{\pi}{1-\pi} \quad \Leftarrow \text{logit transformation}$$

$$\frac{\pi}{1-\pi} \Leftarrow \text{odds}$$

JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING