参考资料
参考资料

Generalized linear model

Linear model

Given an n statistical unit data set

$$\{y_i, x_{i1}, \cdots, x_{ip}\}$$

- y_i is observed values
- $\vec{x}_i$ is a p-vector regressor

Satisfy:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i = \vec{x}_i{}^T \vec{\beta} + \epsilon_i \qquad i = 1, \cdots, n$$

Written into a matrix multiplier:

$$\vec{y} = X\vec{\beta} + \vec{\epsilon}$$

- predicted variables $\vec{y} = [y_1, y_2, \cdots, y_n]^T$
- predictor variables, regressors $X = \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \dots \\ \vec{x}_n^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{ni} & \cdots & x_{np} \end{bmatrix}$
- regression coefficients $\vec{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]^T$
- error term $\vec{\epsilon} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_n]^T$

Note: Predicted variables $\vec{y}$ is different from predicted values are donated as $\hat{\vec{y}}$

Examples:

Consider a situation where a small ball is being tossed up in the air
Height h_i and time t_i (ignoring drag) satisfying:

$$h_i = \beta_1 t_i + \beta_2 t_i^2 + \epsilon_i$$

Model is non-linear with time but linear with $\vec{\beta}$
Here $\vec{x_i} = [t_i, t_i^2]$

Generalized linear model:

In statistics, the generalized linear model (GLM) is a flexible generalization of ordinary linear regression that allows for response variables that have **error distribution models other than a normal distribution**. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a **link function** and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

In a generalized linear model (GLM), each outcome y_i of the dependent variables is assumed to be generated from a particular distribution in an exponential family, a large class of probability distributions that includes the

- Normal
- Binomial
- Poisson
- Gamma distributions, among others.

Exponetial family:

Most of the commonly used distributions form an exponential family or subset of an exponential family, include:

- normal
- exponential
- gamma
- chi-squared
- beta
- Dirichlet
- Bernoulli
- categorical
- Poisson

- Wishart
- inverse Wishart
- geometric

A number of common distributions are exponential families, but only when certain parameters are fixed and known. For example:

- binomial (with fixed number of trials)
- multinomial (with fixed number of trials)
- negative binomial (with fixed number of failures)

Examples of common distributions that are not exponential families are

- Student's t
- Most mixture distributions
- The family of uniform distributions when the bounds are not fixed

pdf:

$$p(x|\theta) = h(x)\exp[\eta(\theta)\cdot T(x) - A(\theta)]$$

Here θ is the parameter of the distribution.

eg

eg

1. The Bernoulli distribution

$$p(x|\pi) = \pi^x(1-\pi)^{1-x} = \exp\{\log(\frac{\pi}{1-\pi})x + \log(1-\pi)\}$$

we have

$$\eta = \frac{\pi}{1-\pi}$$

$$T(x) = x$$

$$A = -\log(1-\pi)$$

$$h(x) = 1$$