Az általánosított lineáris modell (GLM)

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
  1. Van valamilyen eredményváltozó-eloszlás
     - Lineárisnál normális, logisztikusnál Bernoulli
  2. A feltételes várhatóérték valamilyen transzformáltját modellezzük: $g\left[\mathbb{E}\left(Y|\underline{X}\right)\right]=\beta_0+\beta_1 X_1+\ldots+\beta_k X_k$
     - Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szólva annak az inverze)
  3. Elvileg valamit mondani kellhet a varianciáról is
     - Lineárisnál azt, hogy $\sigma_i^2=\sigma_0^2$, logisztikusnál megspóroltuk ezt, mert a várható értéke meghatározta a szórást is (egy paramétere volt az eloszlásnak)

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
- Becslés maximum likelihood-dal
- A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
- Sok minden más is ide tartozik, lássunk még egy példát

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \mathrm{Poi}\left(\lambda\right)$$

$$\log\left[\mathbb{E}\left(Y|\underline{X}\right)\right] = \log\lambda = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$$