MAT-466: Diagnóstico de influencia en GLM II

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Leverage generalizado

Para el modelo de regresión lineal,

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \dots, n,$$

podemos definir el leverage, basado en las siguientes propiedades:

(a) $\mathsf{var}(\widehat{Y}_i) = \sigma^2 h_{ii}$.

(b) $\mathsf{var}(e_i) = \sigma^2(1 - h_{ii})$.

(c) $h_{ii} = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top (\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}})^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})$.

(d) $h_{ij} = \partial \widehat{Y}_i / \partial Y_j$.

Sin embargo, cada una de estas pueden no ser simple de extender para modelos más complejos.

Suponga $\boldsymbol{Y}$ vector n-dimensional con densidad conjunta $f(\boldsymbol{y};\boldsymbol{\theta})$ y $\boldsymbol{\mu} = \mathsf{E}(\boldsymbol{Y})$, tal que $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$. Sea $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{Y})$ y $\widehat{\boldsymbol{Y}} = \boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$.

Wei, Hu y Fung (1998)[1] definieron la matriz de leverage generalizado como:

$$\boldsymbol{GL}(\widehat{\boldsymbol{\theta}}(\boldsymbol{Y})) = \frac{\partial \widehat{\boldsymbol{Y}}}{\partial \boldsymbol{Y}^\top} = \Big\{\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^\top}\,\frac{\partial \widehat{\boldsymbol{\theta}}(\boldsymbol{Y})}{\partial \boldsymbol{Y}^\top}\Big\}\Big|_{\theta=\widehat{\theta}(Y)}.$$

Resultado 1 (Wei, Hu y Fung, 1998):

Suponga que $\boldsymbol{Y}$ tiene log-verosimilitud $\ell(\boldsymbol{\theta};\boldsymbol{y})$ con segunda derivada continua con respecto a $\boldsymbol{\theta}$ y $\boldsymbol{Y}$. Entonces

$$\boldsymbol{GL}(\widehat{\boldsymbol{\theta}}(\boldsymbol{Y})) = \{\boldsymbol{D}_\theta(-\dot{\ell}_{\theta\theta})^{-1}\ddot{\ell}_{\theta Y}\}|_{\theta=\widehat{\theta}(Y)},$$

donde $\boldsymbol{D}_\theta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$, $\ddot{\ell}_{\theta\theta} = \partial^2\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ y $\ddot{\ell}_{\theta Y} = \partial^2\ell(\boldsymbol{\theta};\boldsymbol{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{Y}^\top$.

[1] Scandinavian Journal of Statistics **25**, 25-37.

Leverage generalizado

Para GLMs con ϕ conocido, tenemos

$$\boldsymbol{D}_\beta = \boldsymbol{B}\boldsymbol{X}, \qquad \boldsymbol{B} = \operatorname{diag}(\mathsf{d}\,\mu_1/\,\mathsf{d}\,\eta_1, \dots, \mathsf{d}\,\mu_n/\,\mathsf{d}\,\eta_n),$$

y

$$\ddot{\ell}_{\beta Y} = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{Y}^\top} = \phi \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{B}.$$

Substituyendo $-\ddot{\ell}_{\beta\beta}$ por $\mathsf{E}(-\ddot{\ell}_{\beta\beta})$ obtenemos (aproximadamente)

$$\boldsymbol{GL}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{B}}\boldsymbol{X}(\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{B}}.$$

Así,

$$\boldsymbol{GL}_{ii}(\widehat{\boldsymbol{\beta}}) = \widehat{\omega}_i \boldsymbol{x}_i (\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{x}_i, \qquad \omega_i = (\mathsf{d}\,\mu_i/\,\mathsf{d}\,\eta_i)^2 / V_i.$$

Para el enlace canónico, tenemos

$$\boldsymbol{GL}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{V}}\boldsymbol{X}(\boldsymbol{X}^\top \widehat{\boldsymbol{V}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top.$$

También es posible definir el leverage, notando que

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{W}} \widehat{\boldsymbol{Z}},$$

con $\widehat{\boldsymbol{Z}} = \widehat{\boldsymbol{\eta}} + \widehat{\boldsymbol{W}}^{-1/2} \widehat{\boldsymbol{V}}^{-1/2} (\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})$. Así, la matriz de proyección asociada a la regresión (ponderada) de $\boldsymbol{X}$ contra $\widehat{\boldsymbol{Z}}$ es

$$\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{W}}^{1/2} \boldsymbol{X} (\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{W}}^{1/2}.$$

Es fácil notar que $\widehat{h}_{ii}$ y $GL_{ii}(\widehat{\boldsymbol{\beta}})$ coinciden.

Suponga ϕ conocido. Podemos considerar el desplazamiento de verosimilitudes

$$LD_i = 2\{\ell(\widehat{\boldsymbol{\beta}}) - \ell(\widehat{\boldsymbol{\beta}}_{(i)})\},$$

que no tiene forma explícita.

Sin embargo, podemos usar una aproximación cuadrática

$$LD_i \approx (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \{-\ddot{\ell}_{\beta\beta}(\widehat{\boldsymbol{\beta}})\}^{-1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}).$$

Adicionalmente podemos substituir $-\ddot{\ell}_{\beta\beta}(\widehat{\boldsymbol{\beta}})$ por su valor esperado, obteniendo

$$LD_i \approx \phi(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})^\top \boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}).$$

Diagnóstico por eliminación de casos

Una aproximación de 1-paso para $\widehat{\boldsymbol{\beta}}_{(i)}$ es

$$\widehat{\boldsymbol{\beta}}^1_{(i)} = \widehat{\boldsymbol{\beta}} + \{-\ddot{\ell}_{\beta\beta}(\boldsymbol{\beta})\}^{-1}\boldsymbol{U}_{(i)}(\widehat{\boldsymbol{\beta}}).$$

Substituyendo $-\ddot{\ell}_{\beta\beta}(\widehat{\boldsymbol{\beta}})$ por $\boldsymbol{\mathcal{F}}(\boldsymbol{\beta})$, sigue que

$$\widehat{\boldsymbol{\beta}}^1_{(i)} = \widehat{\boldsymbol{\beta}} - \frac{r_{P_i}\sqrt{\widehat{\omega}_i}}{\phi^{-1/2}(1-\widehat{h}_{ii})}(\boldsymbol{X}^\top\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{x}_i.$$

Esto lleva a la siguiente aproximación

$$LD_i \approx \Big(\frac{\widehat{h}_{ii}}{1-\widehat{h}_{ii}}\Big)t^2_{S_i},$$

con t_{S_i} el residuo estandarizado.

Para motivar ideas considere el modelo perturbado,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathsf{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{D}^{-1}),$$

con $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{\delta})$, con $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ y $\boldsymbol{\delta}_0 = \boldsymbol{1}_n$. Esto lleva a la log-verosimilitud perturbada

$$\begin{aligned} \ell(\boldsymbol{\beta}|\boldsymbol{\delta}) &= -\frac{1}{2}\log|2\pi\sigma^2\boldsymbol{D}^{-1}| - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{D}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= -\frac{n}{2}\log 2\pi\sigma^2 + \frac{1}{2}\sum_{i=1}^n \log\delta_i - \frac{1}{2\sigma^2}\sum_{i=1}^n \delta_i (Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2. \end{aligned}$$

En este caso tenemos

$$-\ddot{\ell}_{\beta\beta} = \frac{1}{\sigma^2}\boldsymbol{X}^\top\boldsymbol{X}, \qquad \boldsymbol{\Delta} = \frac{1}{\sigma^2}\boldsymbol{X}^\top \operatorname{diag}(\boldsymbol{e}),$$

donde $\operatorname{diag}(\boldsymbol{e}) = \operatorname{diag}(e_1, \dots, e_n)$. De este modo, la curvatura normal adopta la forma

$$C_h = \frac{2}{\sigma^2}|\boldsymbol{h}^\top \operatorname{diag}(\boldsymbol{e})\boldsymbol{H}\operatorname{diag}(\boldsymbol{e})\boldsymbol{h}|,$$

con $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. De ahí que $\boldsymbol{h}_{\max}$ es el vector propio dominante de $\boldsymbol{F} = \operatorname{diag}(\boldsymbol{e})\boldsymbol{H}\operatorname{diag}(\boldsymbol{e})$.

Para GLMs considere el esquema de perturbación,

$$\ell(\boldsymbol{\beta}|\boldsymbol{\delta}) = \sum_{i=1}^{n} \delta_i \ell_i(\boldsymbol{\beta}), \qquad 0 \leq \delta_i \leq 1.$$

De este modo,

$$\boldsymbol{\Delta} = \phi \boldsymbol{X}^\top \widehat{\boldsymbol{W}}^{1/2} \operatorname{diag}(\boldsymbol{r}_P), \qquad \operatorname{diag}(\boldsymbol{r}_P) = \operatorname{diag}(r_{P_1}, \dots, r_{P_n}),$$

donde $r_{P_i} = (Y_i - \widehat{\mu}_i)/\sqrt{\widehat{V}_i}$ es el residuo de Pearson. Substituyendo $-\ddot{\ell}_{\beta\beta}$ por $\mathsf{E}(-\ddot{\ell}_{\beta\beta})$, obtenemos

$$C_h = 2|\boldsymbol{h}^\top \operatorname{diag}(\boldsymbol{r}_P) \widehat{\boldsymbol{H}} \operatorname{diag}(\boldsymbol{r}_P) \boldsymbol{h}|,$$

Así, debemos calcular direcciones de interés asociadas a la matriz de curvatura $\boldsymbol{F} = \operatorname{diag}(\boldsymbol{r}_P) \widehat{\boldsymbol{H}} \operatorname{diag}(\boldsymbol{r}_P)$.

Tarea:

- Obtener la matriz de curvatura para el esquema de perturbación, definido por

$$\boldsymbol{Y}(\boldsymbol{\delta}) = \boldsymbol{Y} + \boldsymbol{\delta}.$$

- Llevar a cabo un reporte con el análisis de diagnóstico usando los datos del estudio de la vasoconstricción transitoria en la piel de los dedos (Finney, 1948)[2].

  Estos datos que han sido analizados en Pregibon (1981)[3] y se encuentran disponibles en formato CSV y RDA en el github de la asignatura.

[2] Biometrika **34**, 320-334.

[3] Annals of Statistics **9**, 705-724.