Open access • Journal Article • DOI:10.1111/J.1541-0420.2006.00567.X

Variable selection for logistic regression using a prediction-focused information criterion. — Source link

Gerda Claeskens, Christophe Croux, Johan Van Kerckhoven

Institutions: Katholieke Universiteit Leuven

Published on: 01 Dec 2006 - Biometrics (Blackwell Publishing Inc)

Topics: Focused information criterion, Mean squared error, Feature selection, Estimator and Model selection

Variable Selection for Logistic Regression using a Prediction Focussed Information Criterion

Gerda Claeskens, Christophe Croux and Johan Van Kerckhoven

ORSTAT and University Center for Statistics, K.U. Leuven,

Naamsestraat 69, B-3000 Leuven, Belgium.

email: {gerda.claeskens; christophe.croux; johan.vankerckhoven}@econ.kuleuven.be.

Appendix

Computation of the L_p-norm related risk $r_p(S)$, for p integer.

For $\Lambda_S \sim \mathcal{N}(\lambda, \sigma^2)$, we write $E(|\Lambda_S|^p) = E(|\sigma Z + \lambda|^p)$ where Z has a standard normal distribution. From this it follows that:

$$\begin{aligned} E(|\Lambda_S|^p) &= \frac{1}{\sqrt{2\pi}} \int_{-\frac{\lambda}{\sigma}}^{+\infty} (\sigma z + \lambda)^p \mathrm{e}^{-\frac{z^2}{2}} \, \mathrm{d}z + (-1)^p \int_{-\infty}^{-\frac{\lambda}{\sigma}} (\sigma z + \lambda)^p \mathrm{e}^{-\frac{z^2}{2}} \, \mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{p} \binom{p}{j} \sigma^j \lambda^{p-j} \left\{ \int_{-\frac{\lambda}{\sigma}}^{+\infty} z^j \mathrm{e}^{-\frac{z^2}{2}} \, \mathrm{d}z + (-1)^p \int_{-\infty}^{-\frac{\lambda}{\sigma}} z^j \mathrm{e}^{-\frac{z^2}{2}} \, \mathrm{d}z \right\}. \end{aligned}$$

From this expression, we can derive the following two formulae:

$$E[|\Lambda_S|^p] = \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{p/2} \binom{p}{2j'} 2^{j'} \sigma^{2j'} \lambda^{p-2j'} \Gamma\left(j' + \frac{1}{2}\right),$$

for p even, and

$$\begin{aligned} E[|\Lambda_S|^p] &= \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{(p-1)/2} \binom{p}{2j'} \sigma^{2j'} |\lambda|^{p-2j'} 2^{j'} \Gamma\left(j' + \frac{1}{2}\right) \\ &\quad + \frac{1}{\sqrt{\pi}} \sum_{j=0}^{p} \binom{p}{j} \sigma^j (-|\lambda|)^{p-j} 2^{j/2} \Gamma\left(\frac{j+1}{2}, \frac{\lambda^2}{2\sigma^2}\right), \end{aligned}$$

for p odd. Here, we denoted $\Gamma(\cdot)$ for the gamma function, and $\Gamma(a,x) = \int_x^{+\infty} t^{a-1} \mathrm{e}^{-t} \, \mathrm{d}t$ (for $a > 0$) for the incomplete gamma function.

For p even, say $p = 2r$, the expression can be simplified as follows.

$$
\begin{aligned}
E[|\Lambda_S|^{2r}] &= \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{2r} \binom{2r}{j} \sigma^j \lambda^{2r-j} \int_{-\infty}^{+\infty} z^j \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \\
&= \sqrt{\frac{2}{\pi}} \sum_{j'=0}^{r} \binom{2r}{2j'} \sigma^{j'} \lambda^{2r-2j'} \int_{0}^{+\infty} z^{2j'} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \\
&\overset{u=z^2/2}{=} \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{r} \binom{2r}{2j'} 2^{j'} \sigma^{2j'} \lambda^{2r-2j'} \int_{0}^{+\infty} u^{j'-1/2} \mathrm{e}^{-u} \,\mathrm{d}u \\
&= \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{r} \binom{2r}{2j'} 2^{j'} \sigma^{2j'} \lambda^{2r-2j'} \Gamma\left(j' + \frac{1}{2}\right).
\end{aligned}
$$

For p odd, say $p = 2r + 1$, this leads to

$$
\begin{aligned}
E[|\Lambda_S|^{p}] &= \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{p} \binom{p}{j} \sigma^j \lambda^{p-j} \left\{ \int_{-\frac{\lambda}{\sigma}}^{+\infty} z^j \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z - (-1)^j \int_{\frac{\lambda}{\sigma}}^{+\infty} z^j \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \right\} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{j'=0}^{r} \left\{ \begin{array}{l} \binom{2r+1}{2j'} \sigma^{2j'} \lambda^{2r+1-2j'} \left\{ \int_{-\frac{\lambda}{\sigma}}^{+\infty} z^{2j'} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z - \int_{\frac{\lambda}{\sigma}}^{+\infty} z^{2j'} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \right\} \\ + \binom{2r+1}{2j'+1} \sigma^{2j'+1} \lambda^{2r-2j'} \left\{ \int_{-\frac{\lambda}{\sigma}}^{+\infty} z^{2j'+1} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z + \int_{\frac{\lambda}{\sigma}}^{+\infty} z^{2j'+1} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \right\} \end{array} \right\} \\
&= \sqrt{\frac{2}{\pi}} \sum_{j'=0}^{r} \left\{ \begin{array}{l} \binom{2r+1}{2j'} \sigma^{2j'} \lambda^{2r+1-2j'} \operatorname{sign}(\lambda) \int_{0}^{\frac{|\lambda|}{\sigma}} z^{2j'} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \\ + \binom{2r+1}{2j'+1} \sigma^{2j'+1} \lambda^{2r-2j'} \int_{\frac{|\lambda|}{\sigma}}^{+\infty} z^{2j'+1} \mathrm{e}^{-\frac{z^2}{2}} \,\mathrm{d}z \end{array} \right\} \\
&\overset{u=z^2/2}{=} \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{r} \left\{ \begin{array}{l} \binom{2r+1}{2j'} \sigma^{2j'} \lambda^{2r+1-2j'} \operatorname{sign}(\lambda) 2^{j'} \int_{0}^{\frac{\lambda^2}{2\sigma^2}} u^{j'-\frac{1}{2}} \mathrm{e}^{-u} \,\mathrm{d}u \\ + \binom{2r+1}{2j'+1} \sigma^{2j'+1} \lambda^{2r-2j'} 2^{j'+1/2} \int_{\frac{\lambda^2}{2\sigma^2}}^{+\infty} u^{j} \mathrm{e}^{-u} \,\mathrm{d}u \end{array} \right\} \\
&= \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{r} \left\{ \begin{array}{l} \binom{2r+1}{2j'} \sigma^{2j'} \lambda^{2r+1-2j'} \operatorname{sign}(\lambda) 2^{j'} \{ \Gamma(j' + \frac{1}{2}) - \Gamma(j' + \frac{1}{2}, \frac{\lambda^2}{2\sigma^2}) \} \\ + \binom{2r+1}{2j'+1} \sigma^{2j'+1} \lambda^{2r-2j'} 2^{j'+1/2} \Gamma(j' + 1, \frac{\lambda^2}{2\sigma^2}) \end{array} \right\} \\
&= \frac{1}{\sqrt{\pi}} \sum_{j'=0}^{r} \binom{2r+1}{2j'} \sigma^{2j'} |\lambda|^{2r+1-2j'} 2^{j'} \Gamma\left(j' + \frac{1}{2}\right) \\
&\quad + \frac{1}{\sqrt{\pi}} \sum_{j=0}^{2r+1} \binom{2r+1}{j} \sigma^{j} (-|\lambda|)^{2r+1-j} 2^{j/2} \Gamma\left(\frac{j+1}{2}, \frac{\lambda^2}{2\sigma^2}\right).
\end{aligned}
$$

This ends the proof.