Resampling based testing for linear regression

Logistic regression for binary data.

Let $(y_i, \mathbf{x}_i)_{1\le i\le n}$ be the set of observations where $y_i \in \{0,1\}$. The logistic regression model is,

$$
\begin{aligned}
y_i \sim & \text{Bernoulli}(p_i) \\
p_i = & \frac{1}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}
\end{aligned}
$$

For this problem, $f(\boldsymbol{\beta})$ will be average negative log-likelihood. What is $f'(\boldsymbol{\beta})$?

Likelihood $= P(\mathbf{y} \mid \mathbf{p}) = \prod_{i=1}^n P(y_i \mid p_i) = \prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i}$.

Log-likelihood $= \sum_{i=1}^n [y_i \text{log} p_i + (1-y_i)\text{log}(1-p_i)]$.

Thus, $f(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^n [y_i \text{log} p_i + (1-y_i)\text{log}(1-p_i)]$. This is the loss function or objective function.

$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{1}{n}\sum_{i=1}^n [y_i \frac{\partial \text{log} p_i}{\partial \boldsymbol{\beta}} + (1-y_i)\frac{\partial \text{log}(1-p_i)}{\partial \boldsymbol{\beta}}]$.

We use chain-rule to compute the derivative. $\frac{\partial \text{log} p_i}{\partial \boldsymbol{\beta}} = \frac{\partial \text{log} p_i}{\partial p_i}\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{1}{p_i} \times (-\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i)$

Similarly, $\frac{\partial \text{log}(1-p_i)}{\partial \boldsymbol{\beta}} = \frac{\partial \text{log}(1-p_i)}{\partial p_i}\frac{\partial p_i}{\partial \boldsymbol{\beta}} = (-\frac{1}{1-p_i}) \times (-\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i)$

$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{1}{n}\sum_{i=1}^n [y_i \frac{1}{p_i} \times \{-\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i\} + (1-y_i)(-\frac{1}{1-p_i}) \times \{-\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i\}]$,

which simplifies to

$-\frac{1}{n}\sum_{i=1}^n [-\frac{y_i}{p_i} \times \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i + (\frac{1-y_i}{1-p_i}) \times \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}\mathbf{x}_i] = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2}[\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}]$,

We have $\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}))^2} = p_i(1-p_i)$. Thus, we get $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i p_i(1-p_i)[\frac{y_i(1-p_i)+(1-y_i)p_i}{p_i(1-p_i)}] = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\{y_i(1-p_i)-(1-y_i)p_i\}$.

Finally,

$f'(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\{y_i(1-p_i)-(1-y_i)p_i\} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i(y_i - p_i)$.