Modelos Lineales para Clasificación

Modelos para Multinomial Response

Juan Zamora O.

INSTITUTO DE ESTADÍSTICA

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

Modelos con respuesta categórica

En muchos problemas de regresión, la respuesta está restringida a un conjunto fijo de valores posibles o *categorías de respuesta*

Por ejemplo, la elección de usuarios sobre un conjunto de marcas o los distintos efectos secundarios de un tratamiento médico

Debemos distinguir entre dos condiciones sobre estas categorías:

1. Categoría de respuesta conformada por etiquetas sin ordenamiento inherente
2. Categorias tienen un ordenamiento. Espaciamiento entre categorías no es interpretable

Distribución Multinomial

Representa la generalización de la distribución Binomial

La variable aleatoria multinomial puede tener más de 2 resultados posibles

La variable Y puede tomar cualquier de los valores $1, 2, \ldots, k$ con probabilidades $\pi_1, \pi_2, \ldots, \pi_k$. Es decir, $P(Y = r) = \pi_r$

Distribución Multinomial

Consideremos una muestra de m respuestas. Luego, los componentes del vector $\mathrm{y}^T = (y_1, \ldots, y_{k-1})$ entregan los conteos en cada categoría

Este vector respuesta sigue una distribución multinomial con parámetros m y $\pi^T = (\pi_1, \pi_2, \ldots, \pi_k)$

Este vector tiene función de probabilidad

$$f(y_1, \ldots, y_k) = \frac{m!}{y_1! \ldots y_{k-1}! \, (m - y_1 \ldots - y_{k-1})!} \pi_1^{y_1} \ldots \pi_q^{y_{k-1}} \cdot (1 - \pi_1 - \ldots - \pi_{k-1})^{m - y_1 \ldots - y_{k-1}}$$

Multinomial Logit

Usado para asociar una variable respuesta con categorias no ordenadas a otras variables explicativas

El modelo logit binario tiene la forma

$$log\left(\frac{P(Y=1|\mathrm{x})}{P(Y=2|\mathrm{x})}\right) = \mathrm{x}^T\beta$$

El modelo logit multinomial tiene la misma forma, pero considera $k-1$ logits

$$log\left(\frac{P(Y=r|\mathrm{x})}{P(Y=k|\mathrm{x})}\right) = \mathrm{x}^T\beta_r \ , \ r=1,\ldots,(k-1)$$

k es la categoría de referencia y todas las probabilidades son comparadas con esta última

$$P(Y = k|\mathrm{x}) = \frac{1}{1 + \prod_{s=1}^{k-1} \exp(\mathrm{x}^T \beta_s)}$$

y

$$P(Y = r|\mathrm{x}) = \frac{\exp(\mathrm{x}^T \beta_r)}{1 + \prod_{s=1}^{k-1} \exp(\mathrm{x}^T \beta_s)} \ , \ r = 1, \ldots, (k-1)$$

Interpretación de los parámetros

Analicemos las preferencias de partidos políticos por género

		Partido			
Género	Edad	DC	PS	UDI	RD
H	1	114	224	53	10
	2	134	226	42	9
	3	114	174	23	8
	4	339	414	13	30
M	1	42	161	44	5
	2	88	171	60	10
	3	90	168	31	8
	4	413	375	14	23

Estudiemos el efecto del género $\{1 : \text{M}, 0 : \text{H}\}$ sobre la preferencia política a través de un modelo logit

$$log\left(\frac{P(Y = r|x)}{P(Y = 1|x)}\right) = \beta_{0r} + x_G\beta_r$$

donde $x_G = 1$ cuando responde una mujer y 0 cuando lo hace un hombre. La categoría DC se asocia con el valor 1 y es usada como referencia ($\beta_{01} = 0$)

Luego, los parámetros se interpretan a partir de

$$\beta_{0r} = \log\left(\frac{P(Y = r|x_G = 0)}{P(Y = 1|x_G = 0)}\right)$$

$$\beta_r = \log\left(\frac{P(Y = r|x_G = 1)/P(Y = 1|x_G = 1)}{P(Y = r|x_G = 0)/P(Y = 1|x_G = 0)}\right)$$

$$log\left(\frac{P(Y=2|x)}{P(Y=1|x)}\right) = \beta_{02} + x_G\beta_2$$
$$log\left(\frac{P(Y=3|x)}{P(Y=1|x)}\right) = \beta_{03} + x_G\beta_3$$
$$log\left(\frac{P(Y=4|x)}{P(Y=1|x)}\right) = \beta_{04} + x_G\beta_4$$

Cada vector β_r depende de cada valor r de la respuesta, debido a que la comparación $Y = r$ con $Y = k$ es específica para cara r.

$\exp(\beta_{0r})$ representa los Odds de preferir el partido r en lugar del partido de referencia (DC) por parte de hombres.

$\exp(\beta_r)$ representa la razón que compara los Odds de las preferencias de mujeres con las de los hombres

	β_{0r}	$\exp(\beta_{0r})$	β_r	$\exp(\beta_r)$
DC (1)	0	1	0	1
PS (2)	0.392	1.480	-0.068	0.934
UDI (3)	-1.677	0.187	0.230	1.259
RD (4)	-2.509	0.081	-0.112	0.894

Table: Parámetros estimados para la preferencia de partido en base a género

Para hombres, las Odds de las preferencias de la UDI en lugar de la DC es 0.187. Odds de mujeres vs hombres es 1.259

Medición de la calidad del ajuste

La idea es medir la discrepancia entre las observaciones y el ajuste

Para esto se compara los vectores de observaciones $p_i = y_i/n$ y el vector ajustado π_i, donde ambos vectores tienen dimensión $k-1$.

Deviance

$$D = -2\sum_{r=1}^{k} I(Y_i = r)\log(\widehat{\pi}_{ir})$$

Pearson

$$\chi^2_P = \sum_{i=1}^{n}(p_i - \widehat{\pi}_i)^T \Sigma_i^{-1}(\widehat{\beta})(p_i - \widehat{\pi}_i)$$