2021-10-25

[Μοντελοποίηση Κατηγορικών Ανεξάρτητων Μεταβλητών]

Μοντέλο Παλινδρόμησης Y = εξαρτημένη μεταβλητή

X = κατηγορική ανεξ. μεταβλ.

π.χ. X = κατεύθυνση ανέμου $\in \{N, S, E, W\}$

$X = N$: επίπεδο αναφοράς (αυθαίρετα)

Για τα επίπεδα S, E, W ορίζουμε αντίστοιχες δείκτριες

$$X_S = 1(X=S)$$

$$X_E = 1(X=E)$$

$$X_W = 1(X=W)$$

Y	X	X_S	X_E	X_W
y_1	S	1	0	0
y_2	N	0	0	0
⋮	W	0	0	1
	E	0	1	0
	S	1	0	0
y_N	⋮			

1) Σε κάθε γραμμή το πολύ ένα 1.

2) Αν $X_S = X_E = X_W = 0$ τότε $X = N$

Μοντέλο: $Y = b_0 + b_1 X_S + b_2 X_E + b_3 X_W$

$E(Y) = b_0 + b_1 X_S + b_2 X_E + b_3 X_W$

a) $X = N \Rightarrow X_S = X_E = X_W = 0 \Rightarrow \boxed{E(Y|X=N) = b_0}$

β) $X = S \Rightarrow X_S = 1,\ X_E = X_W = 0 \Rightarrow \boxed{E(Y|X=S) = b_0 + b_1}$

$\Rightarrow \boxed{b_1 = E(Y|X=S) - E(Y|X=N)}$

όμοια)

$$\boxed{\begin{aligned} b_1 &= E(Y|X=S) - E(Y|X=N) \\ b_2 &= E(Y|X=E) - E(Y|X=N) \\ b_3 &= E(Y|X=W) - E(Y|X=N) \end{aligned}}$$

π.χ. ο έλεγχος $t: H_0: b_1 = 0$ vs $H_1: b_1 \neq 0$

Επίσης ο έλεγχος $F: H_0: b_1 = b_2 = b_3 = 0$ vs H_1: τουλάχιστον ένα $\neq 0$

αν απορριφθεί $\Rightarrow \exists$ στατ. σημαντική διαφορά στη $E(Y)$ μεταξύ επ. αναφοράς κ' τουλάχιστον ενός άλλου επιπέδου

$\Rightarrow$ στατ. σημαντική συσχέτιση μεταξύ X κ F.

Logistic Regression

Εξαρτ. μεταβλητή (κατηγορική)

$$G \in \{g_1, \dots, g_K\} \quad (\pi.\chi.\ G \in \underbrace{\{1, 2, \dots, K\}}_{\text{κατηγορίες}})$$

$X =$ διάνυσμα ανεξ. μεταβλητών $= (X_1, \dots, X_p)$

Δείγμα (Training set)

i	G	X_1	$\dots$	X_p	X_0
1	g_1	x_{11}		x_{1p}	1
2	g_2	x_{21}	$\dots$	x_{2p}	1
$\vdots$		$\vdots$			$\vdots$
N	g_N	x_{N1}	$\dots$	x_{Np}	1

$$\log \frac{P(G=k \mid X=x)}{P(G=l \mid X=x)} \begin{cases} \text{αν } > 0 \Rightarrow \hat{f}(x) = k \\ \text{αν } < 0 \Rightarrow \hat{f}(x) = l \end{cases} \quad [k, l \in \{1, \dots, K\}]$$

Στην LDA (γραμμική διαχωριστική ανάλυση)

$$X \mid G=k \sim \mathcal{N}(\mu_k, \Sigma) \quad k = 1, \dots, K$$

κοινό Σ για όλες τις κατηγορίες

προέκυψε $\log \dfrac{P(G=k \mid X=x)}{P(G=l \mid X=x)} = \gamma_0 + \gamma_1 x$ (γραμμικά σύνορα διαχωρισμού)

($f(x) = a + bx$ <u>δεν είναι γραμμική</u> <u>αφφινική</u>)

Στο μοντέλο της λογιστικής παλινδρόμησης

Θεωρούμε την κατηγορία $G=K$ ως κατηγορία αναφοράς

υπόθεση

$$\log \frac{P(G=j \mid X=x)}{P(G=K \mid X=x)} = b_{j0} + b_j^T \cdot x \quad \forall j=1,2,\dots,K-1$$

$b_{j0} \in \mathbb{R}$, $b_j \in \mathbb{R}^p$ άγνωστες παράμετροι

Διάνυσμα παραμέτρων

$$\theta = \{ b_{10}, \underbrace{b_{11}, \dots, b_{1p}}_{b_1}, b_{20}, \underbrace{b_{21}, \dots, b_{2p}}_{b_2}, \dots, b_{K-1,0}, \dots b_{K-1,p} \}$$

$\theta \in \mathbb{R}^{(K-1)(p+1)}$

Αν

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & & \\ 1 & x_{N1} & \dots & x_{Np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix}$$

και

$$b_j = \begin{pmatrix} b_{j0} \\ b_{j1} \\ \vdots \\ b_{jp} \end{pmatrix} \in \mathbb{R}^{p+1}$$

$$\Longrightarrow \log \frac{P(G=j \mid X=j)}{P(G=K \mid X=x)} = x^T b_j = b_j^T x, \quad j=1,\dots,K-1.$$

$$P(G=j \mid X=x) = \underbrace{P(G=K \mid X=x)}_{C(x)} \cdot e^{\beta_j^T x}, \quad j=1,\dots,K-1$$

$$\sum_{j=1}^{K} P(G=j \mid X=x) = C(x) \cdot \left[1 + \sum_{j=1}^{K-1} e^{\beta_j^T x}\right] = 1$$

$$\Rightarrow \quad P(G=K \mid X=x) = \frac{1}{1+\sum_{j=1}^{K-1} e^{\beta_j^T x}}$$

$$P(G=j \mid X=x) = \frac{e^{\beta_j^T x}}{1+\sum_{j=1}^{K-1} e^{\beta_j^T x}}, \quad j=1,\dots,K.$$

Now $\quad P_j(x;\theta) = P(G=j \mid X=x;\theta)$

Fitting (εκτίμηση του θ)

N παρατηρήσεις

i	X	G
1	x_1^T	g_1
$\vdots$	$\vdots$	
N	x_N^T	g_N

Εκτίμηση του θ μέσω μέγιστης πιθανοφάνειας

$$L(\theta; X, G) = \prod_{i=1}^{N} P(G=g_i \mid X = x_i^T; \theta) = \prod_{i=1}^{N} p_{g_i}(x_i;\theta)$$

$$\log L(\theta) = \ell(\theta) = \sum_{i=1}^{N} \log p_{g_i}(x_i;\theta) \leftarrow \max_{\theta}$$

(μέσω αριθμητικών μεθόδων γενικά)

Fitting

N παρατηρήσεις

n	X	G
1	x_1^T	g_1
2	x_2^T	g_2
$\vdots$	$\vdots$	$\vdots$
N	x_N^T	g_N

Εκτίμηση μέσω μέγιστης πιθανοφάνειας

Πιθανοφάνεια

$$- L(\theta; X, G) = \prod_{i=1}^{N} P_{g_i}(\underline{x}_i; \theta)$$

max θ (αριθμητικά)

$$\log L(\theta) = \ell(\theta) = \sum_{i=1}^{N} \log P_{g_i}(\underline{x}_i; \theta)$$

Θα δούμε την περίπτωση $K = 2$

$$\theta = b = (b_{10}, \dots, b_{1p})$$

$$P_1(x; b) = \frac{e^{b^T x}}{1 + e^{b^T x}}$$

$$G \in \{1, 2\} \Rightarrow Y = \begin{cases} 1, & G = 1 \\ 0, & G = 2 \end{cases}$$

$$\ell(\theta) = \sum_{i=1}^{N} \log p_{g_i}(x_i;\theta)$$

Έστω $p(x_i;\theta) = p_1(x_i;\theta)$

$1-p(x_i;\theta) = p_2(x_i;\theta)$

Εδώ $\underline{\log p_{g_i}(x_i;\theta)} = \log p_{y_i}(x_i;\theta) = \begin{cases} \log p(x_i;\theta), & y_i = 1 \\ \log(1-p(x_i;\theta)), & y_i = 0 \end{cases}$

$$= y_i \log p(x_i;\theta) + (1-y_i)\log(1-p_i(x_i;\theta))$$

$$\Rightarrow \underset{\substack{\uparrow \\ \max \theta}}{\ell(\theta)} = \sum_{i=1}^{N}\left[y_i \log p(x_i;\theta) + (1-y_i)\log(1-p(x_i;\theta))\right]$$

Εδώ: $\theta = b = (b_0, \dots, b_p)$, $p(x_i;b) = \dfrac{e^{b^T x_i}}{1+e^{b^T x_i}}$, $1-p = \dfrac{1}{1+e^{b^T x_i}}$

$$\Rightarrow \frac{p(x;b)}{1-p(x;b)} = e^{b^T x} \Rightarrow \log\frac{p}{1-p} = b^T x \Rightarrow$$

$$\left.\begin{aligned} \Rightarrow \log p(x;b) &= b^T x + \log(1-p(x;b)) \\ &= b^T x - \log(1+e^{b^T x}) \end{aligned}\right\}$$

$$\Rightarrow \ell(b) = \sum_{i=1}^{N}\left\{y_i \log p(x_i;b) + (1-y_i)\log(1-p(x_i;b))\right\}$$

$$= \sum_{i=1}^{N}\left\{y_i b^T x_i + y_i \log p(x_i;b) + (1-y_i)\log(1-p(x_i;b))\right\}$$

$$= \sum_{i=1}^{N}\left\{y_i b^T x_i + \log(1-p(x_i;b))\right\}$$

$$\boxed{\begin{aligned} \ell(b) &= \sum_{i=1}^{N}\left[y_i b^T x_i + \log\left(1-p(x_i;b)\right)\right] \\ \ell(b) &= \sum_{i=1}^{N}\left[y_i b^T x_i - \log\left(1+e^{b^T x_i}\right)\right] \end{aligned}} \leftarrow \max_{b}$$

$$\nabla \ell(b) = \begin{pmatrix} \partial \ell / b_0 \\ \vdots \\ \partial \ell / b_p \end{pmatrix}$$

$$\frac{\partial \ell(b)}{\partial b_j} = \sum_{i=1}^{N}\left\{ y_i x_{ij} - x_{ij}\, p(x_i; b) \right\} = \underline{\sum_{i=1}^{N} x_{ij} \overbrace{\left(y_i - p(x_i;b)\right)}^{h_i}}$$

$$\left(\frac{\partial}{\partial b_j} \log\left(1+e^{b x_i}\right) = \frac{\partial}{\partial b_j} \log\left(1+e^{b_0 x_{i0}+\dots+b_j x_{ij}+\dots+b_p x_{ip}}\right) \right.$$

$$\left. = \frac{1}{1+e^{b x_i}} \cdot x_{ij}\, e^{b^T x_i} = x_{ij}\, p(x_i;b) \right)$$

$$\nabla \ell(b) = \begin{pmatrix} \partial \ell / \partial b_0 \\ \vdots \\ \partial \ell / \partial b_p \end{pmatrix} = \begin{array}{c} x_i \\ \begin{pmatrix} \sum_i h_i x_{i0} \\ \vdots \\ \sum_i h_i x_{ip} \end{pmatrix} \end{array} = \sum_{i=1}^{N} \underline{x_i}\, h_i$$

$$\Longrightarrow \boxed{\nabla \ell(b) = \sum_{i=1}^{N} \left(y_i - p(x_i;b)\right) \cdot \underline{x_i}}$$

$\nabla \ell(b) = 0$ (για max $\ell(b)$)

Για λύση του συστήματος χρησιμοποιούμε τη μέθοδο Newton-Raphson ($x_i \in \mathbb{R}^{p+1}$, $b \in \mathbb{R}^{p+1}$)

Στη μονοδ. περίπτωση $f(x) = 0$

μέθοδος Newton: x_0, $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$

Αν $f(x) \overset{(x)}{=} g'(x) = 0 \Rightarrow x_{k+1} = x_k - \dfrac{g'(x_k)}{g''(x_k)}$

Επομένως χρειαζόμαστε τον Hessian matrix του $\ell(b)$

$$\frac{\partial \ell}{\partial b_j \partial b_k} = \frac{\partial}{\partial b_k} \sum_{i=1}^{N} x_{ij} \left(y_i - p(x_i; b) \right) =$$

$$= - \sum_{i=1}^{N} x_{ij} \frac{\partial p(x_i; b)}{\partial b_k}$$

Άσκηση $\dfrac{\partial p(x_i; b)}{\partial b_k} = \dfrac{\partial}{\partial b_k} \left(\dfrac{e^{b^T x}}{1 + e^{b^T x}} \right) = \cdots = x_{ik}\, p(x_i; b)(1 - p(x_i; b))$

$$\Rightarrow \frac{\partial \ell}{\partial b_j \partial b_k} = - \sum_{i=1}^{N} x_{ij} x_{ik}\, p(x_i; b)\left(1 - p(x_i; b)\right)$$

$$\frac{\partial^2 \ell}{\partial b_j \partial b_k} = -\sum_{i=1}^{N} \underbrace{x_{ij}\, x_{ik}}\, p(x_i ; b)\,\big(1 - p(x_i ; b)\big) \Rightarrow$$

$$H_b = \begin{pmatrix} \frac{\partial^2 \ell}{\partial b_0 \partial b_0} & \cdots & \\ & & \\ & \cdots & \frac{\partial^2 \ell}{\partial b_p \partial b_p} \end{pmatrix}$$

$$\boxed{H_b = \sum_{i=1}^{N} \underline{p(x_i ; b)\big(1 - p(x_i ; b)\big)\, x_i x_i^T}}$$

$$\left(x_i x_i^T = \begin{pmatrix} x_0 \\ \vdots \\ x_p \end{pmatrix} (x_0, \ldots x_p) = \begin{pmatrix} x_0^2 & x_0 x_1 & \cdots & x_0 x_p \\ & & & \\ x_p x_0 & \cdots & & x_p^2 \end{pmatrix} \right)$$

$$\frac{\partial \ell}{\partial b} = \nabla \ell(b) = \sum_{i=1}^{N} x_i \left(y_i - p(x_i; b)\right)$$

$$H = -\sum_{i=1}^{N} x_i x_i^T \, \underline{p(x_i; b)\left(1 - p(x_i; b)\right)}$$

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \qquad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix} \qquad b = \begin{pmatrix} b_0 \\ \vdots \\ b_p \end{pmatrix}$$

$$\underline{p}(b) = \begin{bmatrix} p(x_1; b) \\ \vdots \\ p(x_N; b) \end{bmatrix} \qquad W(\underline{b}) = \begin{pmatrix} p(x_1;b)(1-p(x_1;b)) & & 0 \\ & \ddots & \\ 0 & & p(x_N;b)(1-p(x_N;b)) \end{pmatrix}$$

<u>Τότε</u> <u>(δ.ο)</u>:

$$\boxed{\begin{aligned} \nabla \ell(b) &= X^T(y - p) \\ H(b) &= -X^T W X \end{aligned}}$$

<u>Μέθοδος Newton</u> b^0 αυθαίρετο

$$b^{k+1} = b^k - H^{-1}(b^k) \cdot \nabla \ell(b^k)$$

$$\boxed{b^{k+1} = b^k + \underbrace{(X^T W X)^{-1} X^T (y - p)}_{(b_k)}} \qquad k = 0, 1, 2, \ldots$$

Επανάληψη μέχρι να "συγκλίνει"