wo 4 m-logit - lela-gola

multinomial logit.

recap: 2 alternatives.

$y_i = 0$ [base] $y_i = 1$

$$P(y_i = 1 | x_i) = \frac{\exp(\beta_1 + \beta_2 \cdot x_i)}{1 + \exp(\beta_1 + \beta_2 \cdot x_i)} = \frac{\exp(x_i^T \cdot \beta)}{1 + \exp(x_i^T \cdot \beta)}$$

$\exp(0) = 1$

$\exp(x_i^T \cdot 0)$

| view: |

"score"

vector

$$\text{utility}_1(x_i) = x_i^T \cdot \beta^{(1)}$$

$$\text{utility}_0(x_i) = x_i^T \cdot \beta^{(0)}$$

$$\beta^{(0)} = \begin{pmatrix} \beta_1^{(0)} \\ \vdots \\ \beta_k^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$P(y_i = 1 | x_i) = \boxed{\frac{\exp(\text{utility}_1)}{\exp(\text{utility}_0) + \exp(\text{utility}_1)}}$$

$$P(y_i = 0 | x_i) = \frac{\exp(\text{utility}_0)}{\sum_{j=0}^{1} \exp(\text{utility}_j)}$$

a alternatives:

[base] $\text{utility}_0 = x_i^T \cdot \beta^{(0)} = 0$ $\beta^{(0)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

$\text{utility}_1 = x_i^T \cdot \beta^{(1)}$

$\text{utility}_2 = x_i^T \cdot \beta^{(2)}$

$\vdots$

$\text{utility}_{a-1} = x_i^T \cdot \beta^{(a-1)}$

$$P(y_i = j | x_i) = \frac{\exp(\text{utility}_j)}{\sum_{s=0}^{a-1} \exp(\text{utility}_s)}$$ mult. logit

LDA - QDA

linear discriminant analysis — LDA
quadratic — // — QDA

⊕ generative models.

train y, train X → logit ← test x
logit → $\hat{y}$

you can predict y given x

LDA, QDA, MDA ...

* you can predict y given x
* can generate new artificial data (both x and y)

LDA/QDA

$y_i \in \{1, 2, 3 \ldots K\}$ K alternatives

$P(y_i = k) = \pi_k$

$x_i \in \mathbb{R}^n$

$(x_i \mid y_i = k) \sim N(\mu_k, \Sigma_k)$

LDA: each class has its own μ_k
but $\Sigma_k = \Sigma$

QDA: each class has its own μ_k, Σ_k.

ex. $n = 100$ (obs) $y_i \in \{1, 2, 3\}$

$x_i = \begin{pmatrix} a_i \\ b_i \end{pmatrix} \in \underset{\text{free}}{\mathbb{R}^2}$

$2 + 3 + 3 \cdot 2 = 11$ par.

a) how many par-s does LDA model cont-ns?

b) — // — QDA model cont-ns?

$\mu_1 = \begin{pmatrix} \mu_1^a \\ \mu_1^b \end{pmatrix}$

$\vdots$

$\mu_3 = \begin{pmatrix} \mu_3^a \\ \mu_3^b \end{pmatrix}$

a) $\boxed{\pi_1} + \boxed{\pi_2} + \pi_3 = 1$

$\Sigma = \begin{pmatrix} \checkmark & \checkmark \\ & \checkmark \end{pmatrix}$ 3 par-s in Σ

$(x_i \mid y_i = 1) \sim N(\mu_1 ; \Sigma)$

$(x_i \mid y_i = 2) \sim N(\mu_2 ; \Sigma)$

$\boxed{17p}$

ex. Why is it called „linear"

$n=100$

$x_i = \begin{pmatrix} a_i \\ b_i \end{pmatrix} \in \mathbb{R}^2$

$y_i \in \{1,2,3\}$

$$\hat{\mu}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \hat{\mu}_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad \hat{\mu}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix} \qquad \hat{\pi}_1 = 0{,}2 \quad \hat{\pi}_2 = 0{,}3$$

draw a line $P(y_i = 1 \mid x) = P(y_i = 3 \mid x)$

$$P(y_i = 1 \mid x_i) = \frac{P(y_i = 1, x_i)}{f(x_i)} = \frac{P(y_i=1) \cdot f(x_i \mid y_i = 1)}{\sum_{k=1}^{3} P(y_i = k) \cdot f(x_i \mid y_i = k)}$$

$y_i \in \{1,2,3\}$

$x_i \in \mathbb{R}^2$

$$P(y_i = 3 \mid x_i) = \quad \ldots \quad = \frac{P(y_i = 3) \cdot f(x_i \mid y_i = 3)}{\sum P(y_i = k) \cdot f(x_i \mid y_i = k)}$$

$$P(y_i = 1) \cdot f(x_i \mid y_i = 1) = P(y_i = 3) \cdot f(x_i \mid y_i = 3)$$

$$\ln \pi_1 + \ln f(x_i \mid y_i = 1) = \ln \pi_3 + \ln f(x_i \mid y_i = 3)$$

$$\ln 0{,}2 + \ln f_1 = \ln 0.5 + \ln f_3$$

$$(x_i \mid y_i = k) \sim N(\mu_k ; \Sigma)$$

d - dimens of x_i

$d = 2$

$$f(x_i \mid y_i = k) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \cdot \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \cdot \Sigma^{-1} \cdot (x_i - \mu_k)\right)$$

$$\boxed{\ln 0{,}2 - \frac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1} \cdot (x - \mu_1) = \ln 0.5 - \frac{1}{2}(x - \mu_3)^T \cdot \Sigma^{-1} \cdot (x - \mu_3)}$$

$-\frac{1}{2} x^T \cdot \Sigma^{-1} x$ $\qquad$ $-\frac{1}{2} x^T \cdot \Sigma^{-1} x$

$$\boxed{\ln 0.2 - \frac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1} \cdot (x-\mu_1) = \ln 0.5 - \frac{1}{2}(x-\mu_3)^T \cdot \Sigma^{-1} \cdot (x-\mu_3)}$$

$\Sigma^{-1} = S'$ „precision" matrix (inverse of cov)

$$2\ln 0.2 + \underline{x^T \cdot S' \cdot \mu_1} + \underline{\mu_1^T \cdot S' \cdot x} - \mu_1^T \cdot S' \cdot \mu_1 =$$

$$= 2\ln 0.5 + \underline{x^T \cdot S' \cdot \mu_3} + \underline{\mu_3^T \cdot S' \cdot x} - \mu_3^T \cdot S' \cdot \mu_3$$

$x^T \cdot S' \mu_1 = \mu_1^T S' \cdot x \quad \in \mathbb{R}^1$ [scalar]

$$(x_1 \dots x_d) \cdot \left[S' \right] \cdot \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} = (\mu_1 \dots \mu_d) \left[S' \right] \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$S^T = S$ (symmetric)

$$\boxed{\boxed{2(\mu_1^T - \mu_3^T) \cdot S' \cdot} \; x = 2\ln 0.5 - 2\ln 0.2 + \mu_1^T \cdot S' \cdot \mu_1 - \mu_3^T \cdot S' \cdot \mu_3}$$

const

$x = \begin{pmatrix} a \\ b \end{pmatrix}$

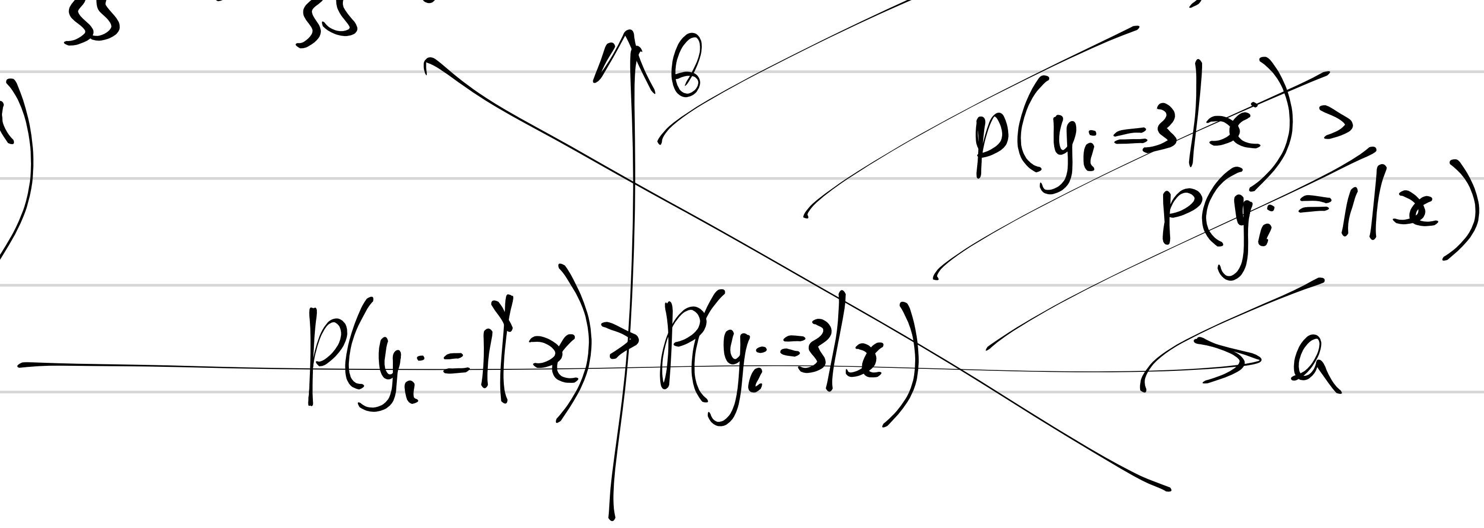
$$2 \cdot \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)^T \cdot \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix}^{-1} = 2 \cdot (2 \ 0) \cdot \frac{1}{35} \begin{pmatrix} 9 & 1 \\ 1 & 4 \end{pmatrix} =$$

$$= \frac{2}{35}(18 \ 2) = \left(\frac{36}{35} \ \frac{4}{35} \right)$$

$$\left(\frac{36}{35} \ \frac{4}{35} \right) \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \text{const}$$

$$\frac{36}{35} a + \frac{4}{35} b = \text{const.}$$

$x = \begin{pmatrix} a \\ b \end{pmatrix}$



exer $\quad n=3 \quad$ discr. $y_i \in \{1,2\}$ $\quad$ cont $x_i \in \mathbb{R}^1$

i	y_i	x_i
1	1	3
2	2	2
3	2	5

estimate LDA model using max. likelihood.

sol-n

$P(y_1=1) \cdot f(x_1=3 \mid y_1=1)$

$$\text{likelihood} = P(y_1=1, x_1=3) \cdot P(y_2=2, x_2=2) \cdot P(y_3=2, x_3=5)$$

not honest probab-s but a mix prob / density.

$$\ln L = \ln \pi_1 + \ln f(x_1=3 \mid y_1=1) + \ln \pi_2 + \ln f(x_2=2 \mid y_2=2) + \ln \pi_2 + \ln f(x_3=5 \mid y_3=2)$$

depends only on π_1

depends only on μ_1, μ_2, σ^2

$\to \max\limits_{\pi_1, \mu_1, \mu_2, \sigma^2}$

$\pi_1 + \pi_2 = 1 \quad$ (1 free par-r)

$(x \mid y=1) \sim N(\mu_1; \sigma^2)$

$(x \mid y=2) \sim N(\mu_2; \sigma^2)$

3 free par-rs: μ_1, μ_2, σ^2

$$\max_{\pi_1} \ \ln \pi_1 + \ln(1-\pi_1) + \ln(1-\pi_1)$$

intuit: $\hat{\pi}_1 = \frac{1}{3}$ (1 obs-n from class 1 and $n=3$)

$$\frac{1}{\hat{\pi}_1} - \frac{1}{1-\hat{\pi}_1} \cdot 2 = 0$$

$$1 - \hat{\pi}_1 = 2\hat{\pi}_1$$

$$3\hat{\pi}_1 = 1$$

$$\hat{\pi}_1 = \frac{1}{3}$$

part II of the log-likelihood function

i	y_i	x_i
1	1	3
2	2	2
3	2	5

$$\ln f(x=3|y=1) + \ln f(x=2|y=2) + \ln f(x=5|y=2)$$

$$\boxed{\ln f(x|y=k) = \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\frac{(x-\mu_k)^2}{2\sigma^2}\right)\right] = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\cdot\ln(\sigma^2) - \frac{1}{2}\frac{(x-\mu_k)^2}{\sigma^2}}$$

part II of $\ln L = -\frac{1}{2}\ln(2\pi)\times 3 - \frac{1}{2}\ln(\sigma^2)\times 3 - \frac{1}{2\sigma^2}\left[(3-\mu_1)^2+(2-\mu_2)^2+(5-\mu_2)^2\right] \to$ max

$$\hat{O}$$

$$\min_{\mu_1,\mu_2,\mu_2} (3-\mu_1)^2+(2-\mu_2)^2+(5-\mu_2)^2 = 0^2+1.5^2+1.5^2 = 2\cdot 2.25 = 4.5$$

$$\hat{\mu}_1 = 3$$

$$\hat{\mu}_2 = \frac{5+2}{2} = 3.5$$

$$\max_{\sigma^2} -\frac{3}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\cdot\boxed{4.5} = h(\sigma^2)$$

$$h'(\sigma^2) = -\frac{3}{2}\cdot\frac{1}{\sigma^2} + \frac{4.5}{2}\cdot\frac{1}{(\sigma^2)^2}$$

$$3\hat{\sigma}^2 = 4.5$$

$$\hat{\sigma}^2 = \frac{4.5}{3}$$

MDA = mixture discr. analysis.

$(x_i|y_i=k) \sim \alpha_1\cdot N(\mu_{k1};\Sigma_{k1}) + \alpha_2\cdot N(\mu_{k2};\Sigma_{k2})$

mixture of normal distributions.