M4: CLASSIFICATION PART B 31.01.2018

Diagnostic paradigm:

$P(Y = k \mid X = x)$ modeled directly

1) Use linear regression: $\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ $Y = \underbrace{X\beta}_{\mu = E(Y \mid X)} + \varepsilon \quad \varepsilon \sim N_n(0, \sigma^2 I)$

$Y \sim N_n(\mu, \sigma^2 I)$ and $\mu = X\beta$

But, if $Y_i \sim bin(1, p_i)$ then $E(Y_i) = p_i$, so why not just do

$$\mu_i = p_i = x_i^T \beta$$

in an ordinary linear regression?

2) Logistic regression: ← a version of binary regression
↑ a special case of Generalized linear models

parameter of interest

a) $Y_i \sim bin(1, p_i)$ random component

b) $\eta_i = \underbrace{x_i^T \beta}_{\beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir}}$ linear predictor (systematic component)

c) link: connecting p_i and $\eta_i = x_i^T \beta$
$[0,1]$ $\mathbb{R}$

Possible links: linear $p_i = x_i^T \beta$

logistic $p_i = \dfrac{\exp(x_i^T \beta)}{1+\exp(x_i^T\beta)} = \dfrac{1}{1-\exp(x_i^T\beta)}$

probit $p_i = \Phi(x_i^T \beta)$

Soon: β's estimated by maximum likehood, but first interpret β.

with regression the goal <u>is</u> interpretation in addition to making a rule for classification

Q: Let $p_i = \dfrac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}$

How can we interpret what happens if x_i increase with one unit? → need odds.

$$\text{Odds} = \frac{p_i}{1-p_i}$$

$p_i = \frac{1}{2}$ $\qquad \frac{\frac{1}{2}}{\frac{1}{2}} = 1$

$p_i = \frac{1}{10}$ $\qquad \frac{\frac{1}{10}}{\frac{9}{10}} = \frac{1}{9} = 0.11$

Why is odds relevant?

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

$$p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \iff p_i + p_i \exp(\eta_i) = \exp(\eta_i)$$

$$\exp(\eta_i) = \frac{p_i}{1-p_i}$$

$$\eta_i = \log\left(\frac{p_i}{1-p_i}\right)$$

$$\eta_i = \text{logit}(p_i)$$

$$\cancel{\log\left(\frac{p_i}{1-p_i}\right)} = \underbrace{\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}\right)}_{\exp}$$

$$\overbrace{\frac{p_i}{1-p_i}}^{} = \exp(\beta_0) \cdot \exp(\beta_1)^{x_{i1}} \cdot \exp(\beta_2)^{x_{i2}} \cdot \exp(\beta_3)^{x_{i3}}$$

$$\frac{P(Y_i = 1 \mid x_i)}{P(Y_i = 0 \mid x_i)}$$

x_{i1}, x_{i2}, x_{i3}

$\rightarrow$ multiplicative model for the odds

So, now ready for increasing x_{i1} to $x_{i1}+1$ – keeping all other – x_i's fixed.

$$\frac{P(Y_i = 1 \mid x_{i1}+1, x_{i2}, x_{i3})}{P(Y_i = 0 \mid x_{i1}+1, x_{i2}, x_{i3})} = \exp(\beta_0) \cdot \exp(\beta_1)^{(x_{i1}+1)} \cdot \exp(\beta_2)^{x_{i2}} \cdot \exp(\beta_3)^{x_{i3}}$$

$$= \exp(\beta_0) \cdot \exp(\beta_1) \cdot \exp(\beta_1)^{x_{i1}} \cdot \exp(\beta_2)^{x_{i2}} \cdot \exp(\beta_3)^{x_{i3}}$$

$$= \frac{P(Y_i = 1 \mid x_{i1}, x_{i2}, x_{i3})}{P(Y_i = 0 \mid x_{i1}, x_{i2}; x_{i3})} \cdot \exp(\beta_1)$$

$\Rightarrow$ odds is multiplid by $\exp(\beta_1)$

You can not use logistic regression if you don't know this!

→ Ex: Default & student

Parameter estimation by maximum likelihood (ML)

scroll to a-b-c = model

$$L(\beta) = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} f(y_i \mid \beta) = \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

$\uparrow$ independent obs

$\downarrow \ln$

$$p_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}$$

$$l(\beta) = \sum_{i=1}^{n} \left(y_i \ln p_i + (1-y_i) \ln(1-p_i) \right)$$

$$= \sum_{i=1}^{n} \left(y_i \underbrace{\ln\left(\frac{p_i}{1-p_i}\right)} + \log(\underbrace{1-p_i}_{\frac{1}{1+\exp(x_i^T \beta)}}) \right)$$

$$= \sum_{i=1}^{n} \left(y_i x_i^T \beta - \ln(1 + \exp(x_i^T \beta)) \right)$$

$$\frac{\partial l}{\partial \beta} \quad [\ (r+1)\times 1 \text{ vector}\ ] \quad = 0$$

vector

→ solve r+1 nonlin eq's

→ not closed form solution

use Newton-Raphson (Fisher scoring)

$\frac{\partial^2 l}{\partial \beta \partial \beta^T}$ Hessian

→ summary printout

Inference: asymptotic theory on ML estimators

N

TMA4295 Stat inf

→ CI & hyp in N-distr.

Evaluation of classifiers

	Pred −	Pred +	total	
True −	TN	FP	N	negative
True +	FN	TP	P	positive
	N^*	P^*		

Sensitivity: $\frac{TP}{P}$ prop. true disease found

syph gitt syph

Specificity: $\frac{TN}{N}$

→ African Heart disease → confusion: 0.5

ROC curve: for all possible cut-offs on classification rule

→ confusion → sens, spec

↓