M4.L2: Classification - Part B

TMA4268
31.01.2019

Diagnostic paradigm:

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

1) Use linear regression: $\underset{n\times 1}{Y} = \underbrace{\underset{(p+1)\times 1}{X\beta}}_{\mu = E(Y|X)} + \varepsilon \quad \varepsilon \sim N_n(0, \sigma^2 I)$

$K = 2$: $Y = 0$ or $Y = 1$

$Y_i \sim \text{binomial}(n=1, p_i)$, $E(Y_i|X_i) = 1 \cdot p_i = p_i$

$\mu_i = E(Y_i \mid X_i) = p_i$ ← use MLR to model that.

2) Let $K = 2$: Logistic regression ← $\underbrace{\text{a version of binary regression}}_{\text{a special case of generalized linear models}}$

$Y_i \in \{0, 1\}$

a) $Y_i \sim \text{bin}(1, p_i)$

parameter of interest.

b) $\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir}$ $\quad r = \#$ covariates

$= x_i^T \beta$

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

Possible link functions:

linear: $p_i = \eta_i = x_i^T \beta$

logistic (sigmoid) $\quad p_i = \dfrac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \dfrac{1}{1 + \exp(-\eta_i)}$

probit $\quad p_i = \Phi(\eta_i)$

Φ: cdf of $N(0,1)$

How to interpret the β's in logistic regression?

With regression the goal is interpretation in addition to making a rule for classification

Let $p_i = \dfrac{e^{\eta_i}}{1+e^{\eta_i}} = \dfrac{e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_r x_{ir}}}{1+e^{\beta_0+\beta_1 x_{i1}+\dots+\beta_r x_{ir}}}$

How can we interpret what happens if x_{i1} increase to $x_{i1}+1$?

$\Rightarrow$ need odds: $\dfrac{p_i}{1-p_i}$ $\qquad p_i = \frac{1}{2} \quad \text{odds} = \dfrac{\frac{1}{2}}{\frac{1}{2}} = 1$

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

$\qquad \vdots$

$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_r x_{ir}$

$p_i = \dfrac{e^{\eta_i}}{1+e^{\eta_i}}, \quad 1-p_i = \dfrac{1+e^{\eta_i}-e^{\eta_i}}{1+e^{\eta_i}} = \dfrac{1}{1+e^{\eta_i}}$

$$\underbrace{\frac{p_i}{1-p_i}}_{\text{odds}} = \frac{e^{\eta_i}/(1+e^{\eta_i})}{1/(1+e^{\eta_i})} = e^{\eta_i} = e^{\beta_0+\beta_1 x_{i1}+\dots}$$

$$\frac{P(Y_i = 1 \mid x_i)}{P(Y_i = 0 \mid x_i)} = \exp(\beta_0) \cdot \exp(\beta_1)^{x_{i1}} \cdots \exp(\beta_r)^{x_{ir}}$$

multiplicative model for the odds

So, what if we increase x_{i1} to $x_{i1}+1$ and keep all other x_i's fixed...

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

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

The odds is multiplied by $\exp(\beta_1)$.

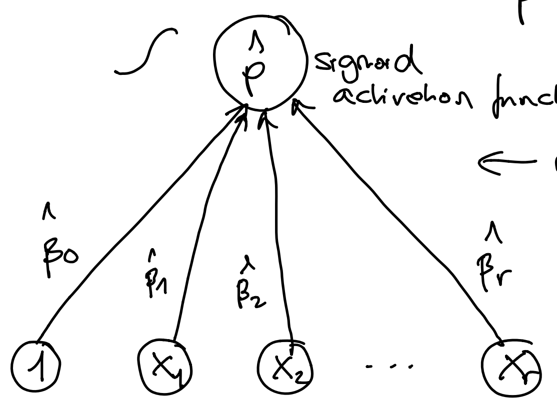
You can't use logistic regression if you don't know this!

Example: $\beta_{student}$?

$\exp(\beta_{student}) = 0.5$ The odds for the student = odds of not student $\cdot$ 0.5

Logistic regression as a network

$$\hat{p} = \frac{e^{\hat{\eta}}}{1+e^{\hat{\eta}}} = \frac{1}{1+e^{-\hat{\eta}}}$$



$\leftarrow \hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_r x_r$

EVALUATING CLASSIFIERS

For $K > 2$ classes $\rightarrow$ confusion matrix and misclassification rate

We focus on $K=2$ classes $\{0, 1\}$

nondisease ÷ disease +

ASSUME:

1) Training data : $\hat{P}(Y=1 \mid X=x) \leftarrow$ build classifier (LDA, QDA, kNN, logistic regr.)

2) Test data: classify to class 1 if $\hat{P}(Y=1 \mid X=x) \geq$ cutoff $\frac{1}{2}$

3) Confusion matrix based on test data

		Predicted		
		$\div$ 0	+ 1	Total
Truth	$\div$ 0	TN	FP	N condition negative
	+ 1	FN	TP	P positive
		N^* negative prediction	P^* positive	

misclassification rate : $\dfrac{FP + FN}{\text{total} \atop N+P}$

Sensitivity : $\dfrac{TP}{P}$

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

a good classifier will have a high sensitivity and a high specificity.

$\Rightarrow$ calculate from table

a) Let $p(x) = \hat{P}(Y=1 \mid X = x)$ when we classify as drawn
if $p(x) \geq 0.5$ $\rightarrow$ sensitivity :
specificity :

b) But, is the cost of misclassification the same for both type of mistakes :
say + when truth is $\div$
say $\div$ +

$\Rightarrow$ instead of putting different costs on this we instead investigate different cut-offs on $p(x)$

$p(x) \geq 0.1 \rightarrow$ classify as disease $\Rightarrow$ calc. sensitivity and specificity

$p(x) \geq 0.2 \rightarrow$ ..

$\vdots$

$p(x) \geq 0.99$

And then I plot x-axis = 1 − specificity

y-axis = sensitivity

Q: $p(x) \geq 0 \Rightarrow$ all $\hat{Y} = 1$

sensitivity: $\frac{TP^{=P}}{P} = 1$

specificity: $\frac{TN^{=0}}{N} = 0$

$p(x) \geq 1 \Rightarrow$ all $\hat{Y} = 0$

sens: $\frac{TP}{P} = 0$

spec: $\frac{TN}{N} = \frac{N}{N} = 1$

What if we just do "random guessing"?

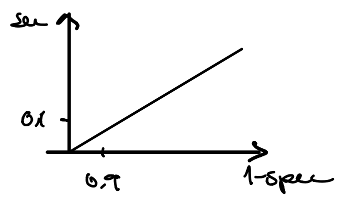
$P(\text{sick}) = 0.1$, $P(\text{healthy}) = 0.9$

* if 0.1 is used as cut-off, for each individual we draw $\hat{Y}$ from $\text{bin}(1, 0.1) \rightarrow \frac{TP}{P} \approx 0.1$, $\frac{TN}{N} \approx 0.9$

* 0.2 $\quad \text{bin}(1, 0.2) \rightarrow \frac{TP}{P} \approx 0.2$, $\frac{TN}{N} \approx 0.8$

This is similar to assign uniformly drawn p's [0,1] to each observation and then setting different cut-offs.

⇒ this will give a ROC curve with AUC = 0.5



This is often used for comparison. An AUC ≈ 0.5 is thus <u>not</u> good.