STA 303/1002-Methods of Data Analysis II

Sections L0101& L0201, Winter 2018

Shivon Sue-Chee

UNIVERSITY OF TORONTO

February 27, 2018

1

$\hat{\pi}_i$ 0

Using Logistic Regression for Classification

Using Logistic Regression for Classification

- Want: predict outcome as

$$y^*|(x_1^*, x_2^*, \ldots, x_p^*) = \begin{cases} 1 \\ 0 \end{cases}$$

- Do: calculate $\widehat{\pi}_M^*$- the estimated probability that $y^* = 1$ based on the fitted model given $X_1 = x_1^*, X_2 = x_2^*, \ldots, X_p = x_p^*$. From this we want to predict that

$$y^* = \begin{cases} 1 & \text{if } \widehat{\pi}_M^* \text{ is large} \\ 0 & \text{if } \widehat{\pi}_M^* \text{ is small} \end{cases}$$

- Need: choose a cut-off probability to distinguish between large and small.

Classification: Approaches to choosing a threshold

Approach 1 - Set cut-off probability as 0.5

- If $\widehat{\pi}_M^* > 0.5$, classify y^* as 1
- Useful if there are equal numbers of 1's and 0's
- Useful if false negatives and false positives are equally bad.

Classification: Approaches to choosing a threshold

$0 < \hat{\pi} < 1$

$\begin{cases} 0 \\ 1 \end{cases}$

Approach 2- Find "best" cut-off probability from data.

- Try different cut-offs and see which gives fewest incorrect classifications
- Useful if proportions of 1's and 0's in data reflect their relative proportions in the population
- Likely to overestimate the proportions of correct predictions that model makes. Then, one should assess model correct classification rates on different data than was used to fit the model.

- More reliable way to find a cut-off prob. is to use Cross-validation.
  - train ← estimates.
  - test ← test estimates.

Confusion Matrix

Observed

type I error

Estimate

	Truth		
Prediction	Positive ($Y = 1$)	Negative ($Y = 0$)	Prop (row)
Positive	TP	FP	PPV=$\frac{TP}{TP+FP}$
Negative	FN type II error	TN	NPV=$\frac{TN}{TN+FN}$
Prop (column)	Sensitivity= TPR=$\frac{TP}{TP+FN}$	Specificity= TNR=$\frac{TN}{TN+FP}$	

- TP: true positive; TN: true negative
- FP: false positive (type I error); FN: false negative (type II error)
- PPV: precision or positive predictive value; false discovery rate=1-PPV
- NPV: negative predictive value; false omission rate=1-NPV

1. Sensitivity (True Positive Rate, TPR)- hit rate
2. Specificity (True Negative Rate, TNR)- prop. of correctly classified negatives
3. False Positive Rate, FPR=1-TNR, fall-out rate
4. False Negative Rate, FNR=1-TPR, miss rate
5. Classification rate=(TN+TP)/(TP+FN+FP+TN); accuracy

Diagnostic Accuracy

- Choose a cut-off probability based on one of the 5 criteria for success of classification that is most important to you.

- High Sensitivity (TPR) makes good screening test.
- High Specificity (TNR) makes a good confirmatory test.
- A screening test followed by a confirmatory test is a good (but expensive) diagnostic procedure.

Confusion Matrix

- From Wikipedia

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive,** Power	**False positive,** Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative,** Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR–}}$	F_1 score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR–) = $\frac{\text{FNR}}{\text{TNR}}$		

https://en.wikipedia.org/wiki/Confusion_matrix

1-way & 2-way ANOVA

Least Squares Regression vs Logistic Regression

$\epsilon_i \sim N(0, \sigma^2)$

Counts

	(Ordinary) Least Squares	(Binomial) Logistic
Response, Y	Normal	# of successes in m trials
Variance	Equal for each level of X	$mp(1-p)$ for each level of X
Model	$\mu_y = \boxed{\beta_0 + \beta_1 X}$	$\log(\frac{\mu}{1-\mu}) = \boxed{\beta_0 + \beta_1 X}$
Model fitting	Least Squares	MLE
Exploratory plot	X vs Y (add line)	logit vs X
Comparing models	Partial F-test AIC/BIC Residuals	LRT/Deviance tests AIC/ BIC (Pearson, Deviance) Residuals
Interpreting	β_1: **change in** μ_y **for unit change in** X	e^{β_1}: **% change in odds for unit change in** X

$\sim Bin(m_i, \pi_i)$.

$m_i \pi_i (1-\pi_i)$.

Wald.

χ^2

Bonferroni vs Tukey's

Tuesday, February 27, 2018 10:48 AM

Bonf: $\left(\bar{y}_i - \bar{y}_j\right) \pm t_{\alpha^*/2,\, df\,\text{Error}} \; S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ $\qquad \alpha^* = \frac{\alpha}{k}$

k = # of pairwise comparisons.

Critical value from Tukey's Range distribution

Tukey's: $\left(\bar{y}_i - \bar{y}_j\right) \pm q^*_{df\,\text{ERROR},\, k} \; S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ $\qquad k = {}^{g}C_2$ all possible pairs.

$S = \sqrt{MSE} = \hat{\sigma} = \sqrt{\hat{\sigma}^2}$ (Residual standard error in R)

Tuesday, February 27, 2018 10:59 AM

How does the estimated logistic model change as we change the "success" and reference level of X?

$\pi_s = P(\text{success})$

$$\text{logit}(\hat{\pi}_s) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_{2,\text{ref}=1}$$

(A) changing "success"

$\pi_f = P(\text{failure})$

$$\text{logit}(\hat{\pi}_f) = \frac{1}{\text{logit}(\hat{\pi}_s)} = \frac{1}{\log\left(\frac{\hat{\pi}_s}{\hat{\pi}_f}\right)} = -\log\left(\frac{\hat{\pi}_s}{\hat{\pi}_f}\right)$$

$$= -\hat{\beta}_0 - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2$$

$$\text{logit}(\hat{\pi}_s) = \log\left(\frac{\hat{\pi}_s}{\hat{\pi}_f}\right) = \frac{\log \hat{\pi}_s}{\log \hat{\pi}_f} = \log \hat{\pi}_s - \log \hat{\pi}_f$$

③ Changing ref. level of X

$$\text{logit}(\hat{\pi}_s) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_{2,\text{ref}1} \qquad \text{ref}1 = 1$$

$$\left[\begin{array}{ll} \text{logit}(\hat{\pi}_s) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 = (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1 & (\text{ref}1 = 1) \\ \text{logit}(\hat{\pi}_s) = \hat{\beta}_0 + \hat{\beta}_1 X_1 & (\text{ref}1 = 0) \end{array}\right.$$

$$\text{logit}(\hat{\pi}_s) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 \mathbb{1}_{\text{ref}1}$$

See Case III — Donner Party Eg.

$$\text{logit}(\hat{\pi}_s) = (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1 - \hat{\beta}_2 \mathbb{1}_{\text{ref}0}$$