Topic 7

1. glm ( ).
2. interpretation of coefficients in logistic model
3. ROC & AUC for logistic model

① glm ( )

lm ( )

glm ($y \sim x_1 + x_2$, data =)

Logistic model :

Response = categorical with 2 categories

→ should be numeric form : 0, 1.

p denotes the probability response = 1.

$\frac{p}{1-p}$ odds of success

⇒ Logistic model :

$$\boxed{\log \frac{p}{1-p}} = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 \cdots$$

churned ~ .

---

② write down fitted model

& interpret coef :

$\hat{p}$ = probability (churned = 1)

$$\log \frac{\hat{p}}{1-\hat{p}} = 3.4 - 0.16 * \text{Age} + 0.07 * I(\text{Married} = 1) + 0.02 * \text{years} + 0.38 * \text{contacts}$$

How to interpret 0.38 (coef of contacts):

when no of contacts increases by 1 then the log-odds of churning increases by 0.38, keeping other variables the same.

$$\log(y_1) = 5 + 6 * x \Rightarrow y_1 = e^{5+6x}$$

$$\log(y_2) = 5 + 6(x+1) \Rightarrow y_2 = e^{5+6(x+1)}$$

$$= 5 + 6x + 6 \qquad y_2 = e^{5+6x+6}$$

$$\frac{y_2}{y_1} = e^{6} \Rightarrow$$

when the no of contacts ↑ by 1, then the $\boxed{\text{odds}}$ of churning changes by $e^{0.38}$ $\boxed{\text{times}}$, keeping other variables the same.

How to interpret 0.07 of Married?

Comparing a married person (Married = 1) vs a non-married when other variables are the same, then the log-odds of churning will be larger by 0.07.

Comparing between a married person vs non-married, then the $\boxed{\text{odds}}$ of churning will change by $e^{0.07}$ $\boxed{\text{times}}$

Column $Pr(>|z|)$ gives p-value for the test:

H_0: variable is NOT significant | → hope to get small

H_1: variable is significant | p-value.

③ ROC & AUC value for a logistic model.

calculate pr(Y = yes)

note: Model M_3 is built using train = full data set.

→ test the goodness of the model by considering test = full data set.

"prediction" from ROCR:

Churned	prob
0	0.08
0	0.1
1	0.8
⋮	⋮

→ tpr, fpr ...

δ 0.05 0.1

threshold

0.1

prob(Y = yes) $\geq$ 0.5 → Yes cate

$\leq$ 0.5 → No cate.

for Churn data: $\boxed{\delta \simeq 0.24}$:

if predicted probability of churning $\geq \delta$ → yes classify as

$< \delta$ → classifying as No.

8000

$\delta = 0.05$ → tpr, fpr.

$\delta = 0.06$ → tpr, fpr

⋮

7300 values of δ