Logistic & Multinomial

Liangjiayi Wang

November 2022

1 Introduction

Rather than modeling this response Y directly, logistic regression models the probability that Y belongs to a particular category.

For example, the probability of default given balance is $Pr(default = Yes|balance)$ or $p(balance)$, will range between 0 and 1. Then for any given value of balance, a prediction can be made for default. For example, one might predict default = Yes for any individual for whom p(balance) > 0.5.

2 Functions

The logistic function will always produce an S-shaped curve of this form, and so regardless of the value of X, we will obtain a sensible prediction.

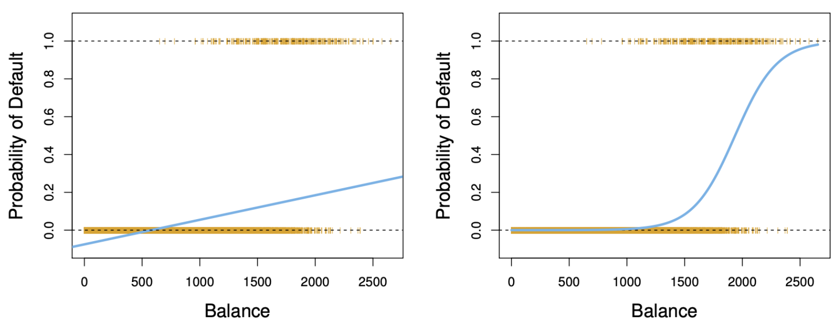


FIGURE 4.2. *Classification using the* `Default` *data.* Left: *Estimated probability of* `default` *using linear regression. Some estimated probabilities are negative! The orange ticks indicate the 0/1 values coded for* `default`*(*`No` *or* `Yes`*).* Right: *Predicted probabilities of* `default` *using logistic regression. All probabilities lie between* 0 *and* 1*.*

logistic function:

$$p(x) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

odds:

$$\frac{p(x)}{1-p(x)} = \frac{P(Y=1|x)}{P(Y=0|x)} = e^{\beta_0+\beta_1 X}$$

which can take from 0 to infinity. Values of the odds close to 0 and ∞ indicate very low and very high probabilities of y, respectively.

By taking the logarithm of both sides, we arrive at log odds or logit:

$$log(\frac{p(x)}{1-p(x)}) = log(\frac{P(Y=1|x)}{P(Y=0|x)}) = \frac{P(Y=1|x)}{P(Y=0|x)} = \beta_0 + \beta_1 X$$

We see that the logistic regression model has a logit that is linear in X.

3 Interpretation

In a logistic regression model, increasing X by one unit changes the log odds by β_1. Equivalently, it multiplies the odds by e^{β_1}

Note: β_1 does not correspond to the change in p(X) associated with a one-unit increase in X. The amount that p(X) changes due to a one-unit change in X depends on the current value of X. But regardless of the value of X, if β_1 is positive then increasing X will be associated with increasing p(X), and if β_1 is negative then increasing X will be associated with decreasing p(X).The fact that there is not a straight-line relationship between p(X) and X, and the fact that the rate of change in p(X) per unit change in X depends on the current value of X.

3.1 X not dummy

	Coefficient	Std. error	z-statistic	p-value
`Intercept`	-10.6513	0.3612	-29.5	<0.0001
`balance`	0.0055	0.0002	24.9	<0.0001

TABLE 4.1. *For the* `Default` *data, estimated coefficients of the logistic regression model that predicts the probability of* `default` *using* `balance`*. A one-unit increase in* `balance` *is associated with an increase in the log odds of* `default` *by* 0.0055 *units.*

Table shows the coefficient estimates and related information that result from fitting a logistic regression model on the Default data in order to predict the probability of default=Yes using balance. We see that $\hat{\beta_1} = 0.0055$; this indicates that an increase in balance is associated with an increase in the probability of default. To be precise, a one-unit increase in balance is associated with an increase in the log odds of default by 0.0055 units.

This null hypothesis implies that $p(x) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$: in other words, that the probability of default does not depend on balance.Since the p-value associated with balance in Table 4.1 is tiny, we can reject H_0. In other words, we conclude that there is indeed an association between balance and probability of default.

3.2 X Dummy 1 for student/ 0 for non-student

	Coefficient	Std. error	z-statistic	p-value
`Intercept`	-3.5041	0.0707	-49.55	<0.0001
`student[Yes]`	0.4049	0.1150	3.52	0.0004

TABLE 4.2. *For the* `Default` *data, estimated coefficients of the logistic regression model that predicts the probability of* `default` *using student status. Student status is encoded as a dummy variable, with a value of* 1 *for a student and a value of* 0 *for a non-student, and represented by the variable* `student[Yes]` *in the table.*

The coefficient associated with the dummy variable is positive, and the associated p-value is statistically significant. This indicates that students tend to have higher default probabilities than non-students:

$$\widehat{\Pr}(\texttt{default}=\texttt{Yes}|\texttt{student}=\texttt{Yes}) = \frac{e^{-3.5041+0.4049\times 1}}{1+e^{-3.5041+0.4049\times 1}} = 0.0431,$$
$$\widehat{\Pr}(\texttt{default}=\texttt{Yes}|\texttt{student}=\texttt{No}) = \frac{e^{-3.5041+0.4049\times 0}}{1+e^{-3.5041+0.4049\times 0}} = 0.0292.$$

$e^{\beta_1} = e^{0.40489} = 1.499$ will be the odds ratio that associates student to the risk of default. This means that:
The student group has a $e^{0.40489} = 1.499$ times the odds of the non-student group of having a default.
or
The student group has 49% $(1.49 - 1 = 0.49)$ more odds of having default than the non-student group.

The coefficient of X means: the log odds of Y is equal to default will increase 0.4049 if a person is student compared to a person is not student, The intercept means the log odds of Y given a person is not a student.Both of estimators are significant.

3.3 0 for student/ 1 for non-student

```
##
## Call:
## glm(formula = default ~ student, family = binomial, data = default.data2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.2970  -0.2970  -0.2434  -0.2434   2.6585
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.09924    0.09071  -34.16  < 2e-16 ***
## student     -0.40489    0.11502   -3.52 0.000431 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2920.6  on 9999  degrees of freedom
## Residual deviance: 2908.7  on 9998  degrees of freedom
## AIC: 2912.7
##
## Number of Fisher Scoring iterations: 6
```

Student status is a dummy variable, with a value of 0 for a student and a value of 1 for a non-student.

$$\hat{Pr}(default = YES|student = no) = \frac{e^{-3.09924-0.40489*1}}{1+e^{-3.09924-0.40489*1}} = 0.0292$$

$$\hat{Pr}(default = YES|student = yes) = \frac{e^{-3.09924-0.40489*0}}{1+e^{-3.09924-0.40489*0}} = 0.0431$$

We see that $\hat{\beta}_0 = -3.09924$, the log-odds of default is -3.09924 for students group.
$\frac{e^{-3.09924-0.40489*0}}{1+e^{-3.09924-0.40489*0}} = 0.0431$ represents that the probability that a student will have a default is 0.0431.

We see that $\hat{\beta}_1 = -0.40489$ and $e^{-0.40489} = 0.667$; this indicates that the non-student group has 33

From above analysis, we find that: $\hat{\beta}_1$ from model $2 = -\hat{\beta}_1 = -0.40489$ from model 1 $\hat{\beta}_0$ from model $2 = \hat{\beta}_0 + \hat{\beta}_1 = -3.50413 + 0.40489 = -0.40489$ from model 1

All the results indicates that non-students tend to have lower default probabilities than students.

4 Lost Sales Example

The p-value of `quote` seems insignificant, so let's build a model without it

```
fit.lost2 <- glm(Status~.,data=df.lost[,-c(1)],family=binomial(logit))
summary(fit.lost2)
```

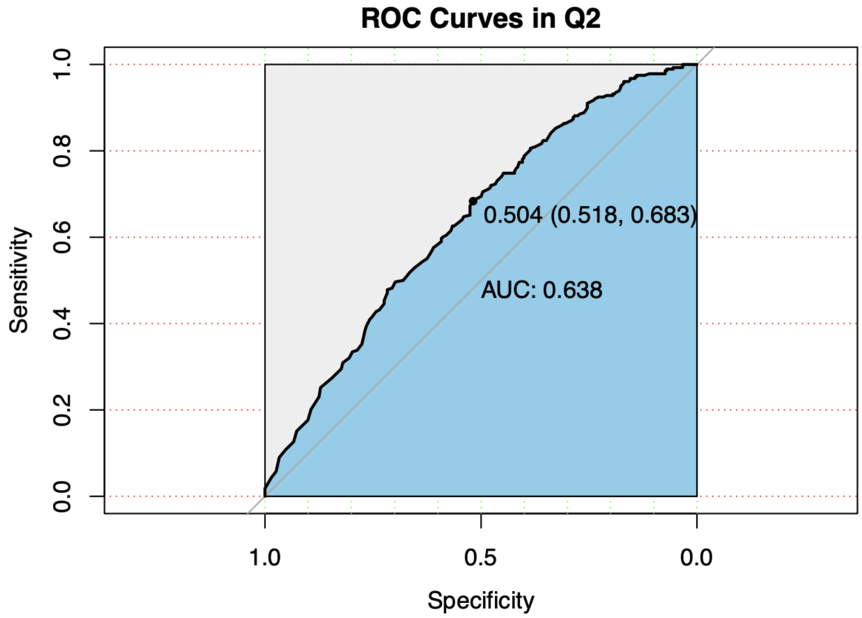
```
##
## Call:
## glm(formula = Status ~ ., family = binomial(logit), data = df.lost[,
##     -c(1)])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4877  -1.1731   0.8893   1.0972   1.9394
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.723499   0.148345   4.877 1.08e-06 ***
## Time.to.Delivery -0.018344   0.003484  -5.266 1.39e-07 ***
## Part.TypeOE      -0.475768   0.196582  -2.420   0.0155 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 762.40  on 549  degrees of freedom
## Residual deviance: 724.22  on 547  degrees of freedom
## AIC: 730.22
##
## Number of Fisher Scoring iterations: 4
```

We use Type I anova to test whether `Quote` is significant.

```
anova(fit.lost1, fit.lost2, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Status ~ Quote + Time.to.Delivery + Part.Type
## Model 2: Status ~ Time.to.Delivery + Part.Type
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       546     723.83
## 2       547     724.22 -1 -0.39542   0.5295
```

We don't have evidence to reject the null at 0.05, the shorter model is adopted.



The data provide some useful information based on the two reasons below:

- Residual deviance is smaller than Null deviance.
- AUC is larger than 0.5 Analysis
- Given ‘Part Type‘ is the same, a small quoted number of calendar days within which the order is to be delivered increases a customer will place an order.
- Given ‘Time to Delivery‘ is the same, quotes with original equipment are less likely to make customers place orders than that with aftermarket.

5 Wine Quality

5.1 Multinominal Logistic Regression

Table 2: Nominal logistic regression

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
Good	(Intercept)	215.13	1.31	164.79	0.00	212.57	217.69
Good	fixed.acidity	0.02	0.08	0.19	0.85	-0.14	0.17
Good	volatile.acidity	-8.95	0.56	-15.97	0.00	-10.05	-7.85
Good	residual.sugar	0.21	0.02	9.50	0.00	0.16	0.25
Good	chlorides	-11.50	3.17	-3.63	0.00	-17.70	-5.29
Good	free.sulfur.dioxide	0.04	0.01	5.93	0.00	0.02	0.05
Good	density	-221.87	1.28	-173.22	0.00	-224.38	-219.36
Good	pH	0.01	0.60	0.02	0.99	-1.16	1.18
Good	sulphates	2.92	0.67	4.39	0.00	1.62	4.23
Good	alcohol	0.94	0.08	11.82	0.00	0.78	1.09
Good	colorwhite	-3.76	0.34	-11.16	0.00	-4.42	-3.10
Just OK	(Intercept)	-175.72	1.18	-148.72	0.00	-178.04	-173.41
Just OK	fixed.acidity	-0.46	0.07	-6.59	0.00	-0.60	-0.32
Just OK	volatile.acidity	-5.42	0.44	-12.28	0.00	-6.28	-4.55
Just OK	residual.sugar	0.00	0.02	-0.02	0.99	-0.04	0.04
Just OK	chlorides	-3.33	2.04	-1.63	0.10	-7.33	0.66
Just OK	free.sulfur.dioxide	0.03	0.01	5.42	0.00	0.02	0.04
Just OK	density	189.71	1.16	163.22	0.00	187.44	191.99
Just OK	pH	-2.52	0.55	-4.59	0.00	-3.59	-1.44
Just OK	sulphates	0.58	0.63	0.92	0.36	-0.65	1.81
Just OK	alcohol	0.46	0.07	6.39	0.00	0.32	0.61
Just OK	colorwhite	-2.80	0.30	-9.43	0.00	-3.38	-2.22

By summarizing the model output, a good wine has below characteristics:

- Good wine should have modest 'fixed.acidity'
- Good wine should have low 'volatile.acidity'
- Good wine should have high 'residual.sugar'
- Good wine should have low 'chlorides'
- Good wine should have high 'free.sulfur.dioxide'.
- Good wine should have modest 'density'
- Good wine should have modest 'PH'
- Good wine should have high 'sulphates'
- Good wine should have high 'alcohol'
- Good wine should tend not to be 'colorwhite'

Here we say a feature is "modest" if the coefficients for the class "Just OK" and the class "Good" have different signs

5.2 Ordinal Logistic Regression

Table 2: Nominal logistic regression

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
Good	(Intercept)	215.13	1.31	164.79	0.00	212.57	217.69
Good	fixed.acidity	0.02	0.08	0.19	0.85	-0.14	0.17
Good	volatile.acidity	-8.95	0.56	-15.97	0.00	-10.05	-7.85
Good	residual.sugar	0.21	0.02	9.50	0.00	0.16	0.25
Good	chlorides	-11.50	3.17	-3.63	0.00	-17.70	-5.29
Good	free.sulfur.dioxide	0.04	0.01	5.93	0.00	0.02	0.05
Good	density	-221.87	1.28	-173.22	0.00	-224.38	-219.36
Good	pH	0.01	0.60	0.02	0.99	-1.16	1.18
Good	sulphates	2.92	0.67	4.39	0.00	1.62	4.23
Good	alcohol	0.94	0.08	11.82	0.00	0.78	1.09
Good	colorwhite	-3.76	0.34	-11.16	0.00	-4.42	-3.10
Just OK	(Intercept)	-175.72	1.18	-148.72	0.00	-178.04	-173.41
Just OK	fixed.acidity	-0.46	0.07	-6.59	0.00	-0.60	-0.32
Just OK	volatile.acidity	-5.42	0.44	-12.28	0.00	-6.28	-4.55
Just OK	residual.sugar	0.00	0.02	-0.02	0.99	-0.04	0.04
Just OK	chlorides	-3.33	2.04	-1.63	0.10	-7.33	0.66
Just OK	free.sulfur.dioxide	0.03	0.01	5.42	0.00	0.02	0.04
Just OK	density	189.71	1.16	163.22	0.00	187.44	191.99
Just OK	pH	-2.52	0.55	-4.59	0.00	-3.59	-1.44
Just OK	sulphates	0.58	0.63	0.92	0.36	-0.65	1.81
Just OK	alcohol	0.46	0.07	6.39	0.00	0.32	0.61
Just OK	colorwhite	-2.80	0.30	-9.43	0.00	-3.38	-2.22

By summarizing the model output, a good wine has below characteristics:

- Good wine should have high ‘fixed.acidity‘.
- Good wine should have low ‘volatile.acidity‘
- Good wine should have high ‘residual.sugar‘
- Good wine should have low ‘chlorides‘.
- Good wine should have high ‘free.sulfur.dioxide‘.
- Good wine should have low ‘density‘.
- Good wine should have high ‘PH‘.
- Good wine should have high ‘sulphates‘
- Good wine should have high ‘alcohol‘.
- Good wine should tend not to be ‘colorwhite‘.

As in proportional odds logistic regression, different levels of ordinal categories share the same coefficients (except intercept), here we only conclude high or low, depending on the coefficient signs. But largely speaking, the conclusions are more or less the same.

Note that the coefficients given by `polr` is negative of the beta coefficients used in class. So from the above `polr` output, for example, `volatile.acidity` has a negative coefficient -3.97. This means

$$\log(\pi_{bad}/(\pi_{ok}+\pi_{good})) = \beta_{01} + ... + \mathbf{3.97} x_{volatile.acidity} + ...,$$

$$\log((\pi_{bad}+\pi_{ok})/\pi_{good}) = \beta_{02} + ... + \mathbf{3.97} x_{volatile.acidity} + ...,$$

both of which means an increase in `volatile.acidity` will lead to an increase in π_{bad} and $\pi_{bad}+\pi_{ok}$. More specifically, one unit increase in `volatile.acidity` gives odds ratio of $e^{3.97}$, or leads to an increase in the odds of $\pi_{bad}/(\pi_{ok}+\pi_{good})$ and $(\pi_{bad}+\pi_{ok})/\pi_{good}$ by a multiplicative factor of $e^{3.97}$. Since both π_{bad} and $\pi_{bad}+\pi_{ok}$ increases as `volatile.acidity` increases, good wine should have low `volatile.acidity`.

more interpretation