CIS 8695

Logistic Regression

Ling Xue

Key Issues

- Logistic regression model formulation
- Logistic regression estimation
- Logit, odds, and probability
- Model fit (Confusion Matrix)
- Variable importance

Logistic Regression

- Extends idea of linear regression to situation where outcome variable is categorical
- Linear regression
  - Predict the value of the continuous Y for a new observation.
- Logistic regression
  - Predict which class a new observation will belong to.

Applications

- Whenever a structured model is needed to explain or predict categorical (in particular, binary) outcome.
- Widely used
  - Classifying customers as returning or nonreturning (classification)
  - Classifying customers as respondents or nonrespondents
  - Predicting the approval or disapproval of a loan based on information such as credit scores
  - Finding factors that differentiate between male and female top executives
- Focus on binary classification

  i.e. Y=0 or Y=1

Two Steps of Logistic Regression

1. Yield estimates of the probabilities of belonging to each class.

    e.g. 0.6

    0.7

    0.3

2. Classify each case into one of the classes based on a cutoff value.

    Suppose the cutoff value is 0.5

    0.6 ->1

    0.7 ->1

    0.3 ->0

The Logit

- Instead of *Y* as outcome variable (like in linear regression), we use a function of Y called the ***logit.***
- Logit can be modeled as a linear function of the predictors.
- The logit can be mapped back to a probability, which, in turn, can be mapped to a class.

Multiple Linear Regression

- Define p = probability of belonging to class 1

- Standard linear function (as shown below) :

$$p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_q x_q$$

q = number of predictors

- Need to relate p to predictors with a function that guarantees $0 \leq p \leq 1$

The Fix

- Assume a nonlinear relationship

logistic response function

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_q x_q)}}$$

- Rewrite

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

The Odds

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

- The odds of an event are defined as:

$$Odds = \frac{p}{1-p}$$

Probability of belonging to class 1

Probability of belonging to class 0

- Or, given the odds of an event, the probability of the event can be computed by:

$$p = \frac{Odds}{1+Odds}$$

Relate Odds to Predictors

$$\log(Odds) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

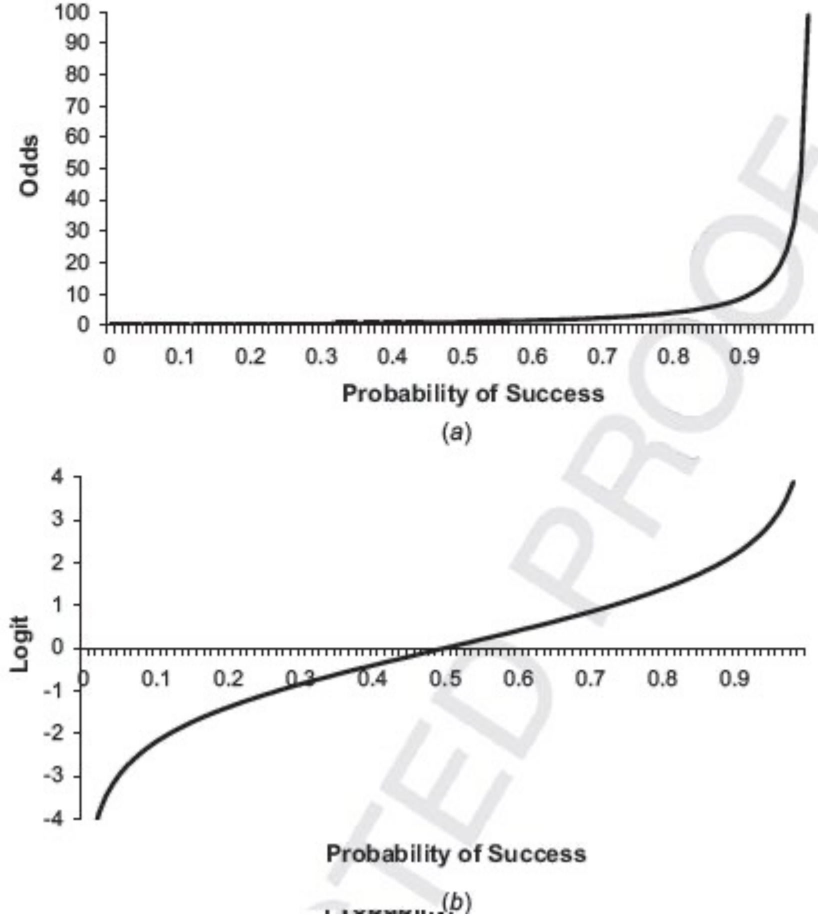
$$\operatorname{logit} = \log(Odds)$$

$$\operatorname{logit}(\mathrm{p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

- So, the logit is a linear function of predictors x1, x2, ...

Takes values from -infinity to +infinity

Odds and Logit as function of *P*



Logistic Regression

- Logistic regression models the log odds of event (or *logit*) as a linear function of the predictors.

$$\operatorname{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

$$\log(Odds) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$$

- It is estimated using **maximum likelihood method** rather than least squares.
  - The maximum likelihood method chooses the parameters to maximize the probability of obtaining the training sample.
- Logistic models are popular for binary responses.

Logistic regression preprocessing

- Similar to linear regression
  - Missing values
  - Categorical inputs
  - Nonlinear transformations of inputs
  - Variable selection (including avoiding multicolinearity)

Example

Personal Loan Offer

- **Goal**: what combination of parameters make a customer more likely to accept a personal loan? (data:UniversalBank)
- **Outcome variable**: accept bank loan (0/1)
- **Predictors:** Demographic info, and info about their bank relationship

ID	Customer ID
Age	Customer's age in completed years
Experience	#years of professional experience
Income	Annual income of the customer ($000)
ZIPCode	Home Address ZIP code.
Family	Family size of the customer
CCAvg	Avg. spending on credit cards per month ($000)
Education	Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional
Mortgage	Value of house mortgage if any. ($000)
Personal Loan	Did this customer accept the personal loan offered in the last campaign?
Securities Account	Does the customer have a securities account with the bank?
CD Account	Does the customer have a certificate of deposit (CD) account with the bank?
Online	Does the customer use internet banking facilities?
CreditCard	Does the customer use a credit card issued by UniversalBank?

Data preprocessing

- Create 0/1 dummy variables for categorical predictors

$$EducProf = \begin{cases} 1 \text{ if education is } \textit{Professional} \\ 0 \text{ otherwise} \end{cases}$$

$$EducGrad = \begin{cases} 1 \text{ if education is at } \textit{Graduate} \text{ level} \\ 0 \text{ otherwise} \end{cases}$$

$$Securities = \begin{cases} 1 \text{ if customer has securities account in bank} \\ 0 \text{ otherwise} \end{cases}$$

$$CD = \begin{cases} 1 \text{ if customer has CD account in bank} \\ 0 \text{ otherwise} \end{cases}$$

$$Online = \begin{cases} 1 \text{ if customer uses online banking} \\ 0 \text{ otherwise} \end{cases}$$

$$CreditCard = \begin{cases} 1 \text{ if customer holds Universal Bank credit card} \\ 0 \text{ otherwise} \end{cases}$$

- Partition 60% training, 40% validation

Last step - classify

- Model produces an estimated probability of being a "1"

- Convert to a classification by establishing cutoff level

- If estimated prob. > cutoff, classify as "1"

Ways to Determine Cutoff

- 0.50 is popular initial choice

- Additional considerations (see Chapter 5)
  - Maximize classification accuracy
  - Maximize sensitivity (subject to min. level of specificity)
  - Minimize false positives (subject to max. false negative rate)
  - Minimize expected cost of misclassification (need to specify costs)

Example, cont’s

- Estimates of β’s are derived through an iterative process called *maximum likelihood estimation*
- Let’s include all 12 predictors in the model now
- Logistic regression output gives coefficients for the logit, as well as odds for the individual terms

Regression Model

Input Variables	Coefficient	Std. Error	Chi2-Statistic	P-Value	Odds	CI Lower	CI Upper
Intercept	-12.497	2.216	31.797	0.0000	0.000	4.85E-08	0.000288
Age	-0.018	0.082	0.050	0.8229	0.982	0.836099	1.152931
Experience	0.017	0.081	0.043	0.8352	1.017	0.867237	1.192774
Income	0.060	0.004	253.440	0.0000	1.062	1.054484	1.07031
Family	0.570	0.095	35.812	0.0000	1.768	1.467059	2.130962
CCAvg	0.149	0.057	6.956	0.0084	1.161	1.039044	1.296827
Mortgage	0.000	0.001	0.010	0.9189	1.000	0.998379	1.001463
Securities A	-0.748	0.373	4.022	0.0449	0.473	0.22773	0.98319
CD Account	3.596	0.443	66.015	0.0000	36.442	15.30749	86.7581
Online	-0.599	0.212	8.015	0.0046	0.549	0.362666	0.831585
CreditCard	-0.785	0.262	8.958	0.0028	0.456	0.272599	0.762542
Education_(	3.984	0.344	133.943	0.0000	53.756	27.37626	105.5537
Education_I	4.172	0.347	144.684	0.0000	64.868	32.86765	128.0227

Estimated Equation for Logit

Input Variables	Coefficient	S
Intercept	-12.497	
Age	-0.018	
Experience	0.017	
Income	0.060	
Family	0.570	
CCAvg	0.149	
Mortgage	0.000	
Securities A	-0.748	
CD Account	3.596	
Online	-0.599	
CreditCard	-0.785	
Education_(	3.984	
Education_I	4.172	

$$Logit = -12.497 - 0.018Age + 0.017Experience + 0.060Income$$
$$+0.570Family + 0.149CCAvg + 0.000Mortgage$$
$$-0.748Securities + 3.596CD - 0.599Online$$
$$-0.785CreditCard + 3.984Edu_{Grad} + 4.172Edu\_Prof$$

$$Odds(Personal\ Loan = Yes) = e^{logit}$$

Converting to Probability

$$p = \frac{Odds}{1 + Odds}$$

Interpreting Odds, Probability

- For predictive classification, we typically use probability with a cutoff value
- For explanatory purposes, odds have a useful interpretation:
  - If we increase x_1 by one unit, holding x_2, x_3 … x_q constant, then
  - e^{β} is the factor by which the odds of belonging to class 1 increase

Evaluating Classification Performance

Performance measures

- Confusion matrix
- % of misclassifications

Validation Data Scoring - Summary Report

Cutoff probability value for success (UPDATABLE)	0.5

Confusion Matrix

	Predicted Class	
Actual Class	1	0
1	119	57
0	19	1805

Error Report

Class	# Cases	# Errors	% Error
1	176	57	32.38636364
0	1824	19	1.041666667
Overall	2000	76	3.8

Performance

Success Class	1
Precision	0.862319
Recall (Sensitivity)	0.676136
Specificity	0.989583
F1-Score	0.757962

Precision and Recall

- Confusion matrix

Actual	Predicted: Positive, 1	Predicted: Negative, 0	
Positive, 1	True Pos. (TP)	False Neg. (FN)	**Recall (*Sensitivity)***
Negative, 0	False Pos. (FP)	True Neg. (TN)	**Specificity**
	Precision		

- Note that all predictions: All = TP + FP + FN + TN
- Some common metrics:
  - **Accuracy** = (TP + TN) / All
  - **Precision** = TP / (TP + FP: predicted positive)
  - **Recall (Sensitivity)** = TP / (TP + FN: actual positive)
  - **Specificity** = TN / (TN + FP)

Multicollinearity

Problem: As in linear regression, if one predictor is a linear combination of other predictor(s), model estimation will fail

- Note that in such a case, we have at least one redundant predictor

Solution: Remove extreme redundancies (by dropping predictors via variable selection – see next, or by data reduction methods such as PCA)

Variable Selection

This is the same issue as in linear regression

- The number of correlated predictors can grow when we create derived variables such as **interaction terms** (e.g. *Income x Family)*, to capture more complex relationships

- Problem: Overly complex models have the danger of overfitting

- Solution: Reduce variables via automated selection of variable subsets (as with linear regression)

Summary

- Logistic regression is similar to linear regression, except that it is used with a categorical response.
- It can be used for explanatory tasks (=profiling) or predictive tasks (=classification).
- The predictors are related to the response Y via a nonlinear function called the *logit.*
- As in linear regression, reducing predictors can be done via variable selection.
- Logistic regression can be generalized to more than two classes.