Conjoint Analysis

Aramayis Dallakyan[1] [2]

[1]Agribusiness teaching Center
Armenia

2019

[2]All errors are my own. armopost@yahoo.com

Outline

Logistic Regression and Conjoint Analysis

- In many situation we deal with the cases when the response variable is *qualitative* or *categorical*.
- In machine and statistical learning literature this cases are called *classification*.
- We are going to discuss one of the most widely-used classifier:*Logistic Regression*.

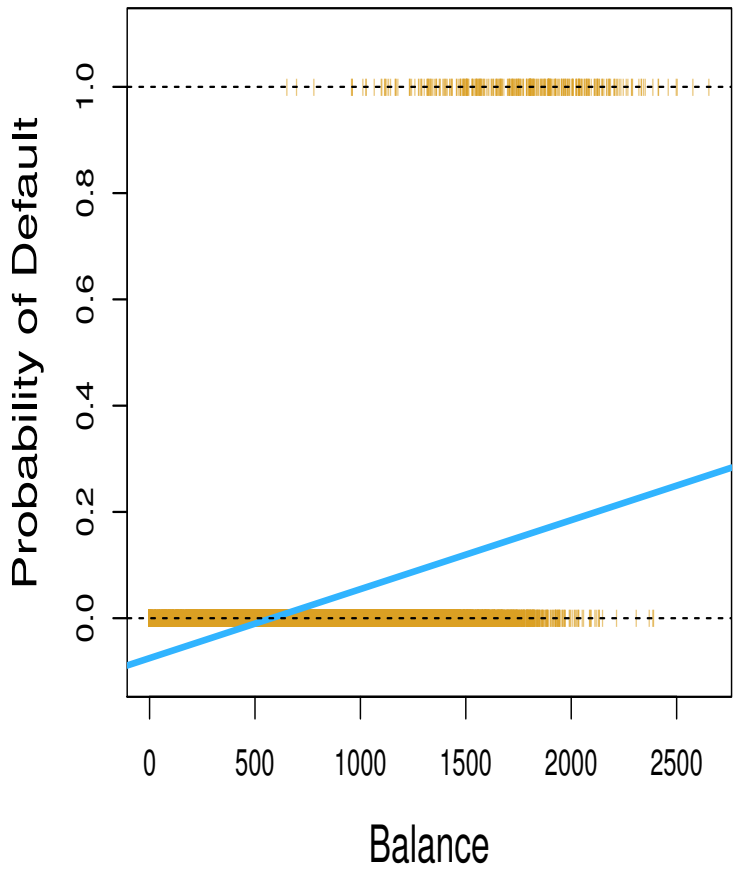
We start by introducing logistic regression. Suppose the response variable is binary (two level). We can code it by 0 and 1.
In classification we are interested on estimating

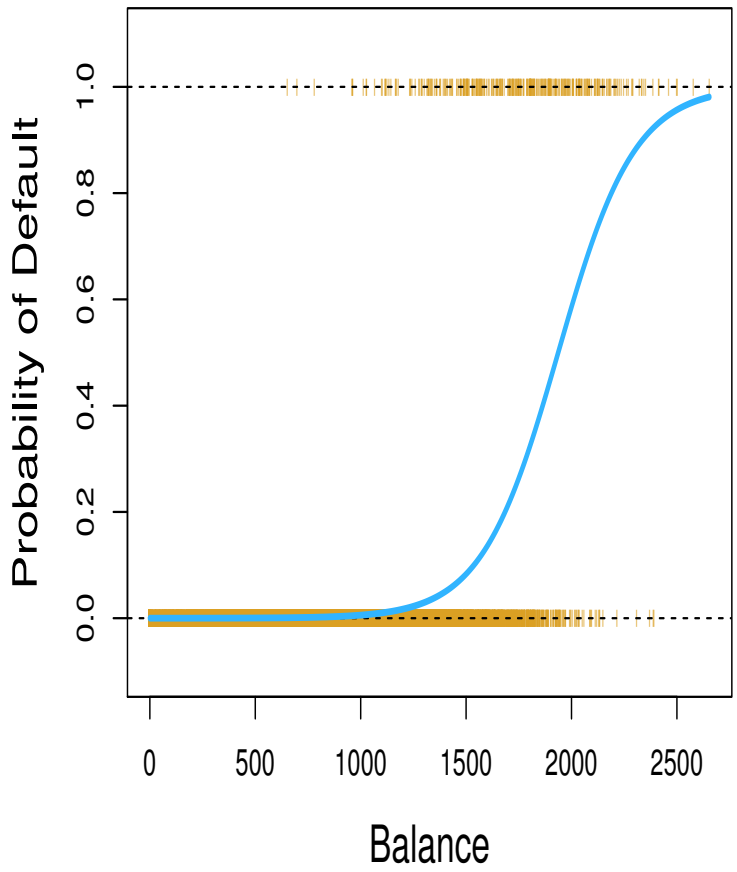
$$p(X) = Pr(Y = 1|X)$$

If we assume linear model than

$$p(X) = \beta_0 + \beta_1 X$$

. The problem is in this cases we may get negative values or in opposite values greater than 1.


Probability of Default
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
Balance


Probability of Default
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
Balance

Thus we need some function which will take X's as a domain and return as arange values from $[0, 1]$. One of such function is, *logistic function*

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

One can show that

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

. This quantity called *odds*. For example if odds of your favorite team to win is 0.9, than it is mean on average nine out of 10 times it will win.

- After taking logarithm from both sides, we get

$$log(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X$$

  The left hand side is called the *log-odds* or*logit*. We can see that logit is linear in X.
- The estimation is done using **Maximum Likelihood(ML)** approach. Think about ML estimation this way, suppose that for each observation ,the probability $y_i = 1$ is $p(X_i)$ (what is probability for $y_i = 0$ and distribution ?).
- Than we want to find β_0 and β_1 that maximize

$$l(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i)^{1-y_i}$$

Suppose an automotive company such as Toyota or Ford is designing a new line of minivans and is trying to determine how large the minivan should be and what type of engine it should have. To inform this decision it would be helpful to understand how customers value those different features.

- Do customers like or dislike hybrid engines?
- If they like them, how much more would they be willing to pay for a hybrid engine?
- Are there segments of customers who like hybrid engines more than other customers?

- In conjoint analysis framework we analyze customers' product choices within a category to understand how features and price affect which product a customer will choose.
- For example, if a customer comes into the store and purchases a 30 oz. jar of Hellman's brand canola mayonnaise for $3.98, we can conceptualize this as the customer choosing that particular type of mayonnaise among all the other mayonnaise available at that store.
- This data on customers' choices can be analyzed to determine which features of a product (e.g., package size, brand, or flavor) are most attractive to customers and how they trade off desirable features against price.

- The idea of determining the features of a product may sound familiar from the Lecture Note 1.
- However here customer makes a *choice* among several options, each of which has its own set of attributes.
- To solve the issue, we are going to adapt multinomial logit model, the most frequently used choice model in marketing.

Choice-based **conjoint analysis**, is a survey method where customers are asked to make choices among products with varying features and prices.Conjoint surveys give marketers information about how customers choose products by asking respondents to answer survey questions.

Which of the following minivans would you buy?

Assume all three minivans are identical other than the features listed below.

	Option 1	Option 2	Option 3
	6 passengers 2 ft. cargo area gas engine $35,000	8 passengers 3 ft. cargo area hybrid engine $30,000	6 passengers 3 ft. cargo area gas engine $30,000
I prefer (check one):	☐	☐	☑

In this question, respondents are asked to choose from three product profiles, each with a

- specific passenger capacity
- cargo capacity
- engine type
- price.

The product options in the survey are called *alternatives* and the product features are called *attributes*. This conjoint analysis study has three alternatives in each question, described by four attributes. Each attribute occurs at some level. For example, the possible levels for cargo capacity in our example survey are 2 ft. and 3 ft.

We analyze the data from a conjoint survey ,where

- 200 respondents
- each answer 15 questions
- each question offers 3 alternatives
- so so each respondent sees a total of $15 \times 3 = 45$ product profiles

Our job is to find out the important attributes based on the choice.