Logistic Regression

Part 2

Logistic Regression Matrix

Female	Young	Music	Sport	News	Clicked	Forecast
1	1	0.2	0.5	0.8	1	x= $\beta_0 + \beta_F \cdot 1 + \beta_Y \cdot 1 + \beta_M \cdot 0.2 + \beta_S \cdot 0.5 + \beta_N \cdot 0.8$ $p = \frac{e^x}{1 + e^x}$
0	1	0.4	0.3	0.5	0	As above, but with these attribute values.
1	0	0.8	0.1	0.9	0	…
…	…	…	…	…	…	…

Each row describes an user with 5 attributes, 2 binary, 3 real.

The sixth column *Clicked* is the outcome.

The column *Forecast* is the expected value *p*.

We maximize the total fitness, i.e. we minimize the total error over all rows.

A real life example in marketing

A telco wants to launch a new service, say home Internet connection.

Before starting to sell in the whole market, we try a sample of customers in order to build a predictive model. The model will forecast the probability that a customer of a certain profile will agree to purchase the new service.

If a customer purchases the service, we get a profit of 2,000€.

Trying to sell is costly: we can contact a single customer via telemarketing or a population of customers via television advertising.

A customer profile has 3 attributes, which we use as predictors: Education, Mobility, Income.

We make each attribute binary with choices like:

- Education = 1 if the customer has a degree, 0 otherwise.
- Mobility = 1 if the customer change home more than one time in 10 years.
- Income = 1 if the customer has income greater than 50,000€ per year.

The answer, i.e. the variable to predict, is *Purchases*, i.e. the customer purchases the service.

Each cell contains the ratio customers purchasing / contacted

	Low education		High education	
	Low mobility	High mobility	Low mobility	High mobility
Low income	153/2160 = 0.071	226/1137 = 0.199	61/886 = 0.069	233/1091 = 0.214
High income	147/1363 = 0.108	139/547 = 0.254	287/1925 = 0.149	382/1415 = 0.270

The model

$$Prob(Purchases = 1\ |Education = edu, Mobility = mob, Income = inc\ ) = \frac{exp(x)}{1 + exp(x)}$$

where

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

and *edu*, *mob*, *inc* are all binary values, zero or one. Say 1 means High and 0 means Low, but the choice of coding scheme for values is arbitrary.

With some algebraic manipulation, we get

$$\beta_0 = log\left(\frac{Prob(Purchases = 1|edu = 0, mob = 0, inc = 0)}{Prob(Purchases = 0|edu = 0, mob = 0, inc = 0)}\right)$$

The quantity in parenthesis is the *odds*, i.e. the ratio between probability of success and probability of failure.

$$Prob(Purchases = 1 \,|Education = edu, Mobility = mob, Income = inc) = \frac{exp(x)}{1 + exp(x)}$$

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

$$\beta_0 = log\left(\frac{Prob(Purchases = 1|edu = 0, mob = 0, inc = 0)}{Prob(Purchases = 0|edu = 0, mob = 0, inc = 0)}\right)$$

The odds are another way to express probability.

Say prob(success) = 25% and prob(failure) = 75%.

The odds is 1/3. This means that in a fair bet your stake is 1/3 of your opponent's one. They bet 3€, you bet 1€.

In our context the odds are the "right" investment you can do to try to get a potential customer to buy.

You can spend this amount of money contacting a single customer in phone, or many on television. In any case, you can invest at most 1€ for each 3€ a customer can bring if purchasing, because you will succeed in converting 1/4 of the customers.

$$Prob(Purchases = 1\ |Education = edu, Mobility = mob, Income = inc) = \frac{exp(x)}{1 + exp(x)}$$

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

$$\beta_0 = log\left(\frac{Prob(Purchases = 1|edu = 0, mob = 0, inc = 0)}{Prob(Purchases = 0|edu = 0, mob = 0, inc = 0)}\right)$$

If the logistic regression tools, after examining historical data, estimates the probability of selling is 25%, then it gives you a value $\beta_0 = \log\left(\frac{1}{3}\right)$.

This is the intuitive interpretation of the first parameter: the log of the odds for customers with profile (0, 0, 0).

These customers are the baseline market segment.

The parameter β_0 allows you to compute the odds, and the probability, for these segment.

$$Prob(Purchases = 1 \mid Education = edu, Mobility = mob, Income = inc) = \frac{exp(x)}{1 + exp(x)}$$

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

$$\beta_1 = log\left(\frac{Odds(edu = 1, mob = 0, inc = 0)}{Odds(edu = 0, mob = 0, inc = 0)}\right)$$

The parameter $\beta_1$is of a different nature.

It is the logarithm of an odds-ratio (instead of an odds).

Its interpretations is: (the log of) the variation in the right investment to substitute a customer of the baseline segment with a customer of the segment which differs from the baseline only because a value 1 in the factor *Education*.

Parameters β_2 and β_3 have the same meaning, related to factors *Mobility* and *Income*, respectively.

$$Prob(Purchases = 1 \mid Education = edu, Mobility = mob, Income = inc) = \frac{exp(x)}{1 + exp(x)}$$

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

$$\beta_1 = log\left(\frac{Odds(edu = 1, mob = 0, inc = 0)}{Odds(edu = 0, mob = 0, inc = 0)}\right)$$

Each parameter, except the first, is about variation.

Now, you can estimate the investment to exchange a customer for another.

What does it mean "exchange"?

You could spend in order to leave a customer for another.

Or you can spend in order to exchange the customers attitude.

How? With advertising.

So, logistic regression helps you to decide the right investment in advertising.

$$Prob(Purchases = 1 \mid Education = edu, Mobility = mob, Income = inc) = \frac{exp(x)}{1 + exp(x)}$$

$$x = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot mob + \beta_3 \cdot inc$$

$$\beta_1 = log\left(\frac{Odds(edu = 1, mob = 0, inc = 0)}{Odds(edu = 0, mob = 0, inc = 0)}\right)$$

The logistic regression equation gives each factor a certain importance (the related parameter) measuring its impact on the probability of success, if moving form level 0 to level 1.

Note that we have only terms of first degree, i.e. a weighted average of single factors. We do not have joint factors, nothing like ... $+ \beta_{12} \cdot edu \cdot mob$

We are assuming the interactions among factors are not relevant.

Once again, we are assuming *independence* among factors.