Advanced Statistical Modelling: Logistic Regression

Ricard Meyerhofer & Joel Cantero

4/11/2019

Exploratory data analysis

As explained in the problem statement, our dataset is composed by 28645 calls from JYB data that contain the following attributes:

Variable	Description	Attribute type
id	Customer ID	Client
age	age in years	Client
job	(admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown)	Client
marital	Marital status (Divorced, married, single, unknown)	Client
education	Education level (basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown) (No, yes, unknown)	Client
default	is he/she a defaulter?	Client
housing	does he/she has a mortgage? (No, yes, unknown)	Client
loan	does he/she has a personal loan? (No, yes, unknown)	Client
contact	phone type (cellular, telephone)	Call
month	month of the call	Call
day_of_week	day of the call (mon, tue, wed, thu, fri)	Call
campaign	does he/she has a personal loan? (No, yes, unknown)	Campaign
pdays	does he/she has a personal loan? (No, yes, unknown)	Campaign
previous	does he/she has a personal loan? (No, yes, unknown)	Campaign
poutcome	does he/she has a personal loan? (No, yes, unknown)	Campaign
emp.var.rate	employment variation rate (quarterly)	Indicators
cons.price.idx	Consumer Price Index (monthly)	Indicators
cons.conf.idx	Consumer confidence index (monthly)	Indicators
euribor3m	euribor a 3 mesos (daily)	Indicators
nr.employed	number of employed (quarterly)	Indicators
Y	The customer subscribed the deposit? (yes,no)	Response

With the original variables, fit the complete model without interactions and using the logit link function

aaaaa

Evaluate possible first order interactions(between two factors or between a factor and a covariable) and include them in the model (if there were any)

aaaaa

Perform an automatic variables selection based on the AIC and BIC. Make a comparasion of the models and argue which one is chosen.

aaaaa

Validate the model y checking the assumptions

aaaaa

Interpret the final model

aaaaa