ASM Homework 2

Generalized Linear Model for JYB data

Maria Gkotsopoulou & Ricard Monge Calvo & Amalia Vradi

22/10/2019

Exploratory Data Analysis

We check for any missing values or attributes without a value and find none nor NAs.

Table 1: Numerical variables

variable	class	min	mean	median	max
age	integer	17.00	39.98	38.00	98.00
campaign	integer	1.00	2.56	2.00	43.00
pdays	integer	0.00	962.63	999.00	999.00
previous	integer	0.00	0.17	0.00	7.00
emp.var.rate	numeric	-3.40	0.08	1.10	1.40
cons.price.idx	numeric	92.20	93.58	93.80	94.77
cons.conf.idx	numeric	-50.80	-40.48	-41.80	-26.90
euribor3m	numeric	0.63	3.62	4.86	5.04
nr.employed	numeric	4963.60	5167.00	5191.00	5228.10

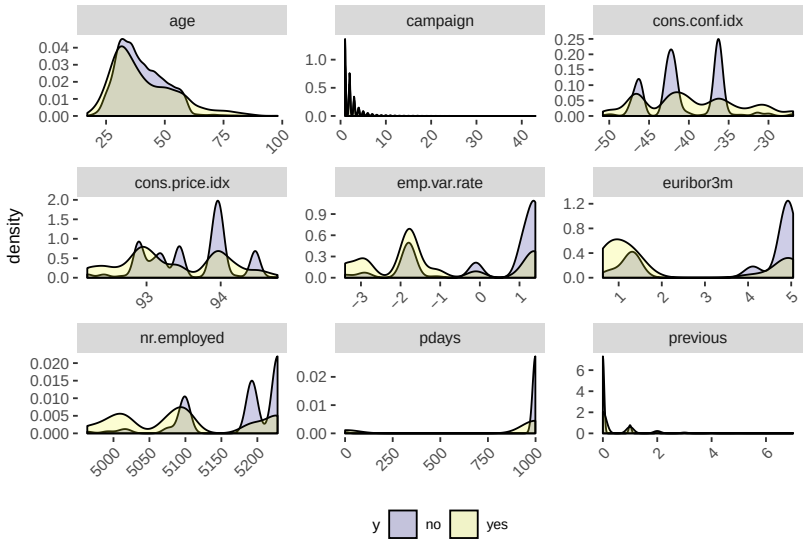
Table 2: Categorical variables

attribute	# levels
job	12
marital	4
education	8
default	3
housing	3
loan	3
contact	2
month	10
day_of_week	5
poutcome	3
y	2

attribute	level_1	level_2	level_3	level_4	level_5	level_6	level_7	level_8	level_9	level_10	level_11	level_12
job	admin.	blue-collar	entrepreneur	housemaid	management	retired	self-employed	services	student	technician	unemployed	unknown
marital	divorced	married	single	unknown								
education	basic.4y	basic.6y	basic.9y	high.school	illiterate	professional.course	university.degree	unknown				
default	no	unknown	yes									
housing	no	unknown	yes									
loan	no	unknown	yes									
contact	cellular	telephone										
month	apr	aug	dec	jul	jun	mar	may	nov	oct	sep		
day_of_week	fri	mon	thu	tue	wed							
poutcome	failure	nonexistent	success									
y	no	yes										

We are interested in predicting whether the customer subscribed to the deposit, so our target variable y is a binary one.

We now look closer into the relation between y and all the numerical variables.



Complete Model

Evaluation First Order Interactions

Automatic Variable Selection process

Model comparison

Model validation

Model interpretation