MKT 500T HW 1

Sami Cheong

Jan 28 2018

Projecting customer retention rates

Data

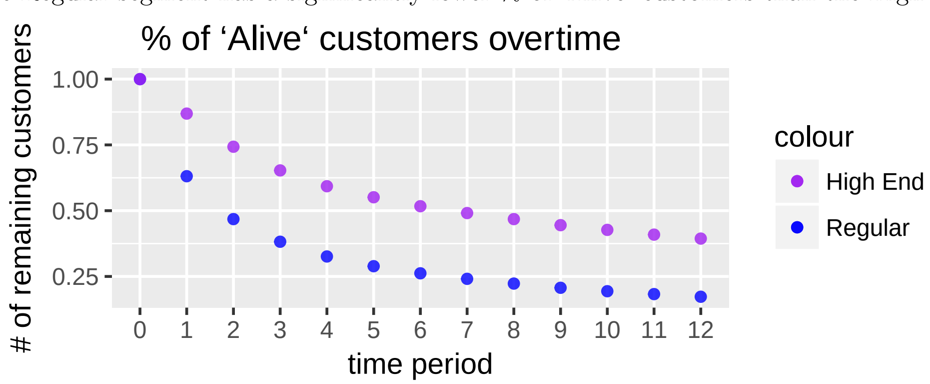
We are given the following customer retention data,starting with 1000 customers, but gradually decreasing over 12 years. For this assignment, we will use only data from year 1-4:

Year	Regular	High_End
0	1000	1000
1	631	869
2	468	743
3	382	653
4	326	593

and the goal is to predict the following retention rates in the following periods:

Year	Regular	High_End
5	289	551
6	262	517
7	241	491
8	223	468
9	207	445
10	194	427
11	183	409
12	173	394

Notice that the Regular segment has a significantly lower % of 'Alive' customers than the High End semengt:



Basic statistics

First, let's take a look at the churn rates and retention rates of each customer segment:

```
data[,'Churn_R']<- c(0, -diff(data$Regular))
data[,'Churn_H']<- c(0, -diff(data$High_End))
data[,'Ret_R']<- data$Regular/1000
data[,'Ret_H']<- data$High_End/1000
insamp.data<-data[1:sample.size,]
kable(insamp.data)
```

Year	Regular	High_End	Churn_R	Churn_H	Ret_R	Ret_H
0	1000	1000	0	0	1.000	1.000
1	631	869	369	131	0.631	0.869
2	468	743	163	126	0.468	0.743
3	382	653	86	90	0.382	0.653
4	326	593	56	60	0.326	0.593

Model

Based on this training data, we are interested in knowing the period t in which a customer churns $P(T = t)$, with the following assumption:

- θ is the probability that a customer witll 'flip' and switch brands.
- $P(T = t|\theta)$ is the probability that a customer 'flip' at time t given θ, which follows the geometric distribution with density $f(t|\theta) = \theta(1 - \theta)^t$.
- θ follows a density function $g(\theta)$ characterized by $\Gamma(\alpha, \beta)$.

Now, since $P(T = t) = \int_0^1 f(t|\theta)p(\theta)d\theta$, after some algebra we can deduce that $P(T = t)$ follows the shifted Beta Geometric Distribution, which is defined by the following:

$$P(T = t) = \begin{cases} \frac{\alpha}{\alpha+\beta} & \text{for } t = 1, \\ \frac{\beta+t-1}{\alpha+\beta+t-1} P(T = t - 1) & \text{for } t = 2, 3, 4, \ldots \end{cases}.$$

In here, α and β are obtained by maximizing the log-likelihood function defined by

$$l(\alpha, \beta) = \sum_{t=1}^{4} c_t \ln(P(T = t|\theta)) + n_4 ln(P > 4|\theta),$$

where c_i is the number of customers who churns at time t and n_4 is the number of 'survivors' at the end of the 4^{th} season.

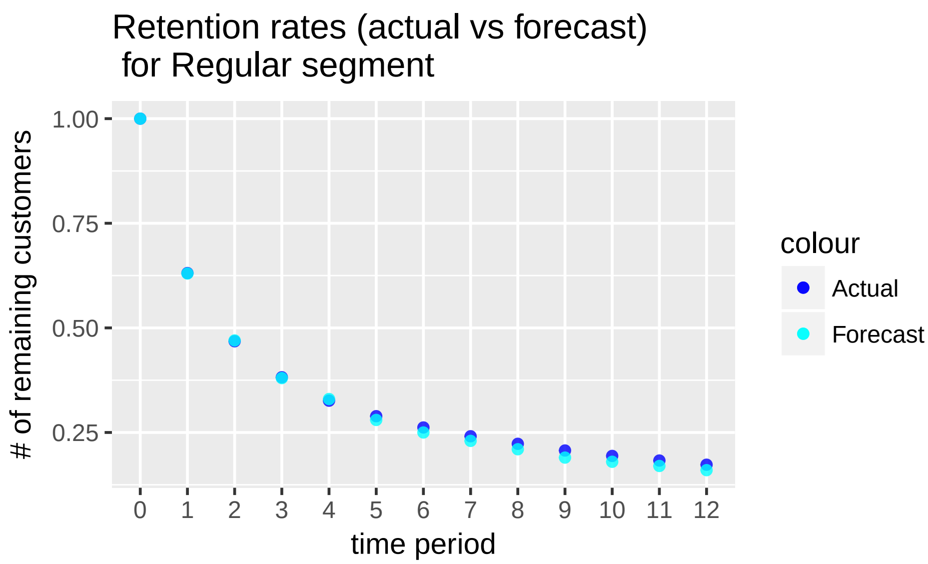
This parameter estimation procedure is implemented in R using the code provided in `sBG.r`:

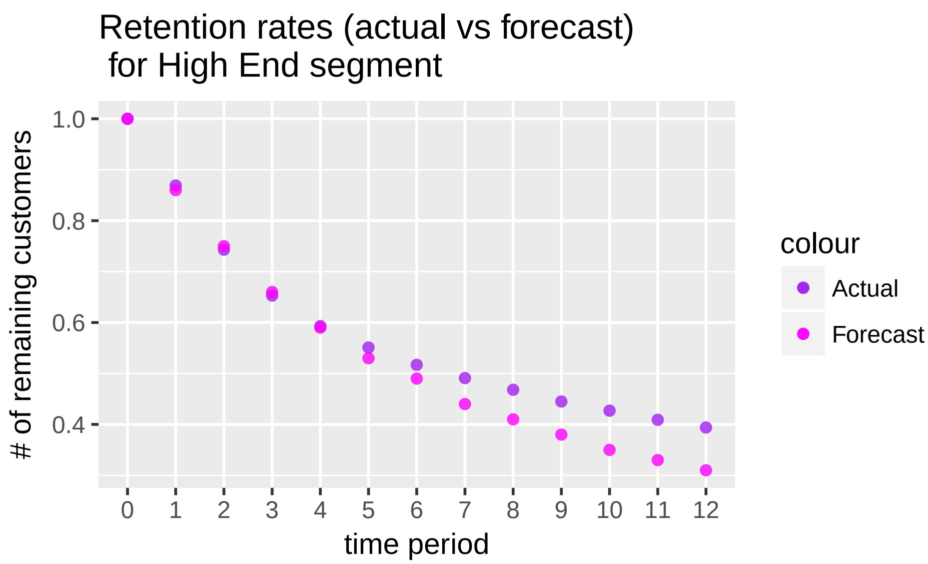
Result

Using 4 years of training data, we obtained the following best estimates for α and β. Since both

	alpha	beta
Regular	0.763665	1.295830
High End	1.280976	7.790384

Putting it all together:





Year	Ret_R	Forecast_R	APE_R	Ret_H	Forecast_H	APE_H
1	0.631	0.63	0.00	0.869	0.86	0.01
2	0.468	0.47	0.00	0.743	0.75	0.01
3	0.382	0.38	0.01	0.653	0.66	0.01
4	0.326	0.33	0.01	0.593	0.59	0.01
5	0.289	0.28	0.03	0.551	0.53	0.04
6	0.262	0.25	0.05	0.517	0.49	0.05
7	0.241	0.23	0.05	0.491	0.44	0.10
8	0.223	0.21	0.06	0.468	0.41	0.12
9	0.207	0.19	0.08	0.445	0.38	0.15
10	0.194	0.18	0.07	0.427	0.35	0.18
11	0.183	0.17	0.07	0.409	0.33	0.19
12	0.173	0.16	0.08	0.394	0.31	0.21

Conclusion

In here, we calculate the in and out-of-sample average percent error. While both in-sample MAPE is relatively small (0.5 % and 1% respecitvely), we see that the out-of-sample MAPE get a lot higher (6% and 13% respecitvely). Interestingly, when we compare the MAPE using 4 years of data vs that trained from 7 years of data, we see a better in-sample MAPE. This suggests that 4 years of data provide a better fit. However, the out-of-sample MAPE did a lot worse when using the model trained by 4 years of data, so even though it has a good fit, it is not as predictive compared to the model that uses 7 years of data.

	Regular	High End	High End (7 years data)
insample.MAPE	0.00500	0.01	0.017
outsample.MAPE	0.06125	0.13	0.028

Another interesting point, we observed that comparing to the example (where $\alpha = 0.668$ and $\beta = 3.806$) done in class, both α and β here are estimated to be higher, which changes the density quite a bit as it movies the curve for $g(\theta)$ to an entirely different quadrant.