MKT 500T

Spring 2018 HW1

Sami Cheong

Feb 04 2018

Projecting customer retention rates

Data

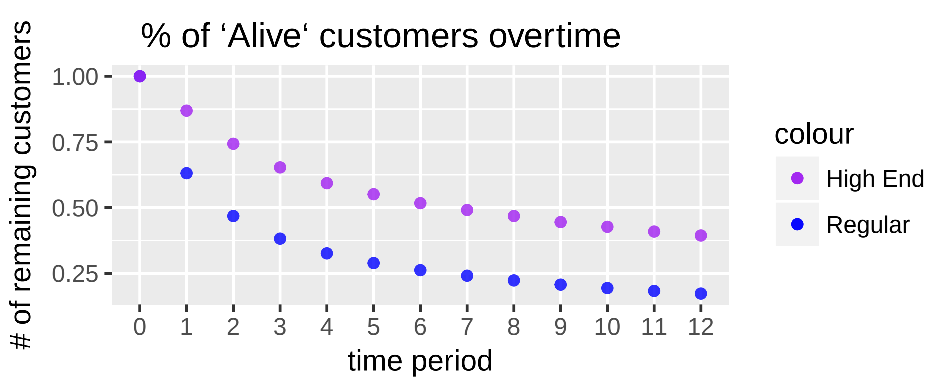
We are given the following customer retention data,starting with 1000 customers, but gradually decreasing over 12 years. For this assignment, we will use data from years 1-4:

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

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 assumptions:

- 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 $\Gamma(\alpha, \beta)$.

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`, presented as follows:

```
# define 'survivors' for each segment
surv.R<-insamp.data$Regular[nrow(insamp.data)]
surv.H<-insamp.data$High_End[nrow(insamp.data)]

par.init=c(0.6,0.6)
# get optimized parameter values for Regular and High End segments:
result.reg = optim(par=par.init,fn=sBG.LL,
                   churn.data = insamp.data$Churn_R[2:nrow(insamp.data)],
                   survivors = surv.R,control=list(fnscale=-1,reltol=1e-18))
par.final.reg = exp(result.reg$par)
```

```
result.high = optim(par=par.init,fn=sBG.LL,
                    churn.data = insamp.data$Churn_H[2:nrow(insamp.data)],
                    survivors = surv.H,control=list(fnscale=-1,reltol=1e-18))
par.final.high = exp(result.high$par)

# compute density function values:
pmf.reg<-sBG.pmf(par.final.reg[1],par.final.reg[2],12)


pmf.high<-sBG.pmf(par.final.high[1],par.final.high[2],12)

# compute log-likelihood valuse:
LL.reg<-sBG.LL(par.final.reg,insamp.data$Churn_R[2:nrow(insamp.data)],
               surv.R,debug.flag=F)

LL.high<-sBG.LL(par.final.high,insamp.data$Churn_H[2:nrow(insamp.data)],
                surv.H,debug.flag=F)

# collect values:
LL<-c(LL.reg,LL.high)
```

Result

Using 4 years of training data, we obtained the following best estimates for α and β. Notice that estimates for the High End segment has a worse likelihood value than the Regular segment, this suggests that the High End model may not have as good a fit compared to the Regular one.

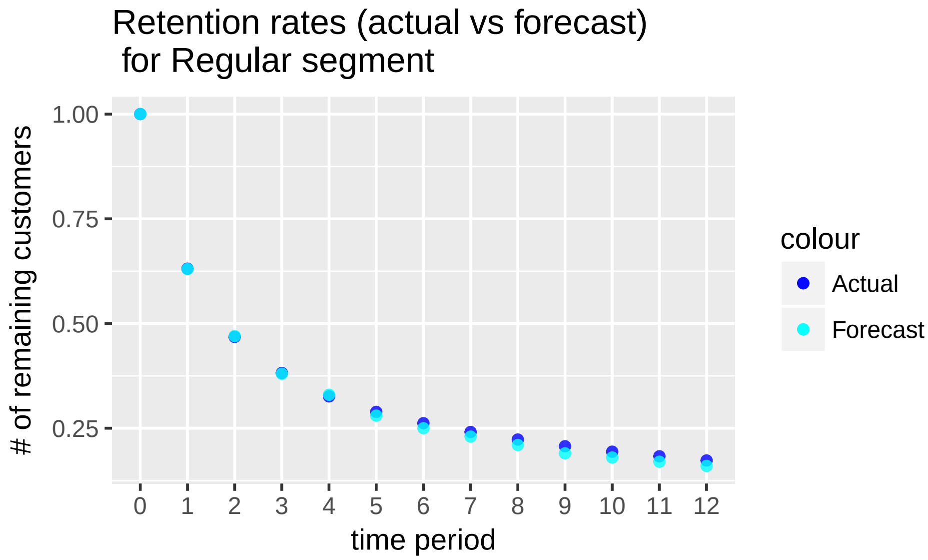
	alpha	beta	log-likelihood
Regular	0.763665	1.295830	-1427.933
High End	1.280976	7.790384	-2654.386

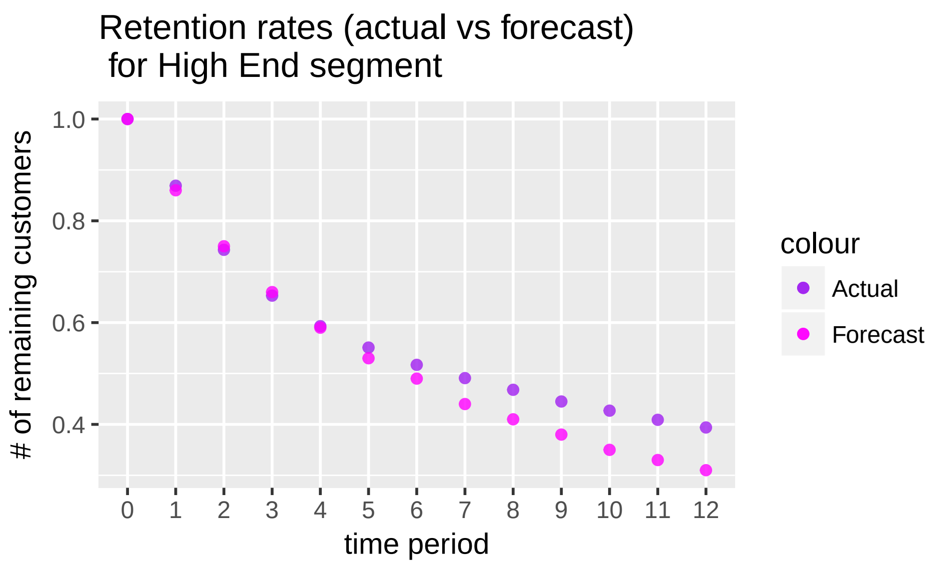
Putting it all together:

```
data[,'Forecast_R']<-round(c(1,1-cumsum(pmf.reg)),2)
data[,'Forecast_H']<-round(c(1,1-cumsum(pmf.high)),2)

data[,'APE_R']<-round(abs(data$Ret_R-data$Forecast_R)/data$Ret_R,2)

data[,'APE_H']<-round(abs(data$Ret_H-data$Forecast_H)/data$Ret_H,2)
```





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

Comparing model trained on 4 years vs 7 years of data:

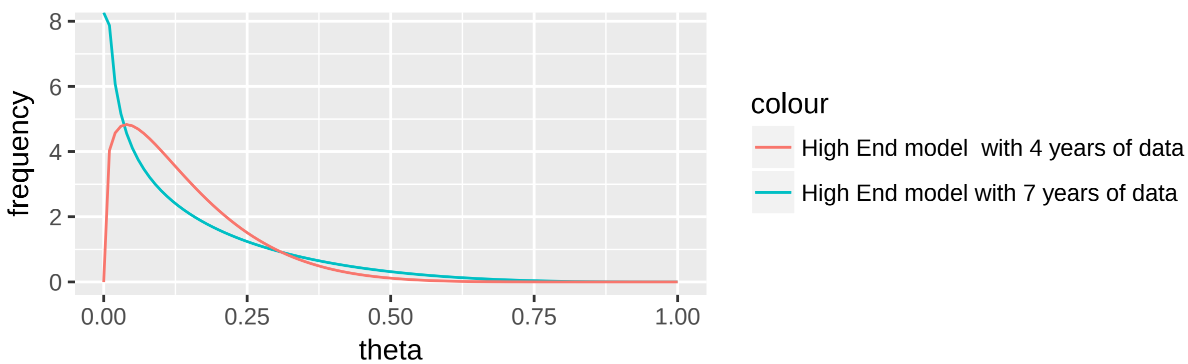
In here, we calculate the in and out-of-sample average percent error for model trained with 4 years. While both in-sample MAPE is relatively small (0.5 % and 1% respecitvely), the out-of-sample MAPE's are much higher (6% and 13% respecitvely). Repeating the same procedure using 7 years of data instead gives the following:

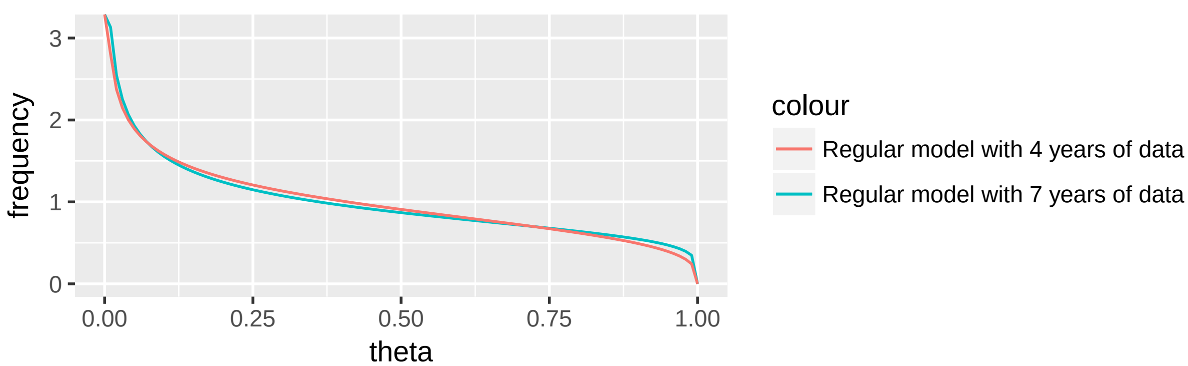
	Regular	High End	Regular (7 years data)	High End (7 years data)
insample.MAPE	0.00500	0.01	0.011	0.017
outsample.MAPE	0.06125	0.13	0.015	0.028

Interestingly, MAPE's from models trained with 4 years of data were actually lower when we looking at in-sample data. This suggests that they provide a better fit for the data, but ultimately the models are not as predictive compared to those that used 7 years of data.

How did the training data change the underlying distribution for $g(\theta)$?

Another interesting observation is that, comparing to the example for High End data set (where $\alpha = 0.668$ and $\beta = 3.806$), both α and β here are estimated to be higher when using less traiing data. This changes the density quite a bit as it moves the curve for $g(\theta)$ to an entirely different quadrant, linking that to the MAPE, we see that the discrepensies between the High End models trianed are much higher compared to those using the data from the Regular segment.





Conclusion

In this assignment, we investigated the following:

- Analyzed two sets of customer churn data over 12 years
- Applied the shifted Beta-Geometric distribution to model the probability of churn (θ)
- Estimated the parameters (α, β) that govern θ using maximum-likelihood estimation
- Compared model performance using 4 years of trianing data vs 7 years of training data