

ISLR::Caravan

Sahba Salarian

Feb. 2019

#Introduction

##DATA EXPLANATION

###Attribute Information

`summary`(Caravan)

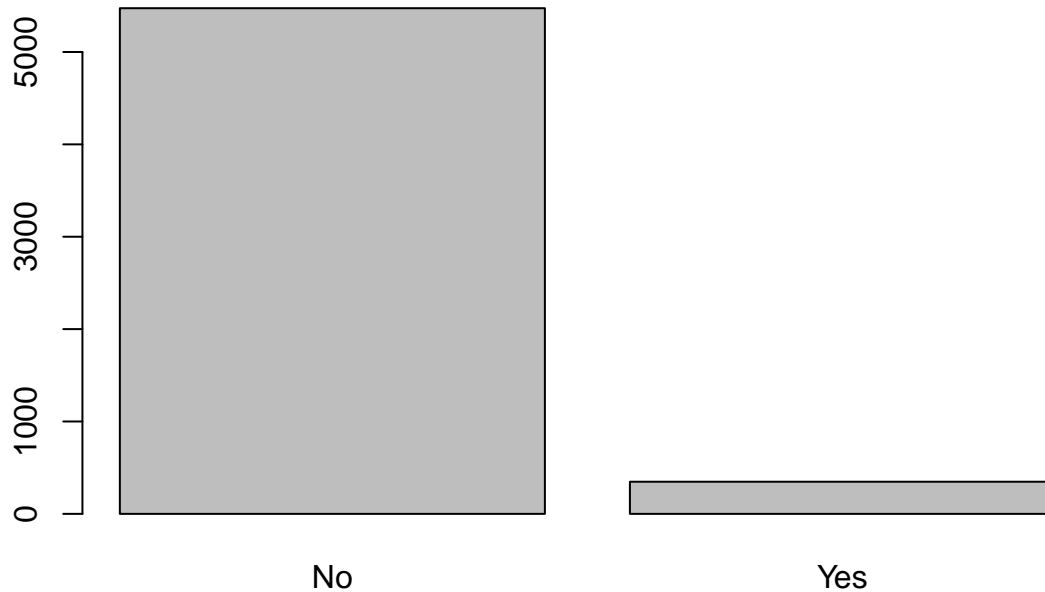
##	MOSTYPE	MAANTHUI	MGEMOMV	MGEMLEEF
##	Min. : 1.00	Min. : 1.000	Min. :1.000	Min. :1.000
##	1st Qu.:10.00	1st Qu.: 1.000	1st Qu.:2.000	1st Qu.:2.000
##	Median :30.00	Median : 1.000	Median :3.000	Median :3.000
##	Mean :24.25	Mean : 1.111	Mean :2.679	Mean :2.991
##	3rd Qu.:35.00	3rd Qu.: 1.000	3rd Qu.:3.000	3rd Qu.:3.000
##	Max. :41.00	Max. :10.000	Max. :5.000	Max. :6.000
##	MOSHOOFD	MGODRK	MGODPR	MGODOV
##	Min. : 1.000	Min. :0.0000	Min. :0.000	Min. :0.00
##	1st Qu.: 3.000	1st Qu.:0.0000	1st Qu.:4.000	1st Qu.:0.00
##	Median : 7.000	Median :0.0000	Median :5.000	Median :1.00
##	Mean : 5.774	Mean :0.6965	Mean :4.627	Mean :1.07
##	3rd Qu.: 8.000	3rd Qu.:1.0000	3rd Qu.:6.000	3rd Qu.:2.00
##	Max. :10.000	Max. :9.0000	Max. :9.000	Max. :5.00
##	MGODGE	MRELGE	MRELSA	MRELOV
##	Min. :0.000	Min. :0.000	Min. :0.0000	Min. :0.00
##	1st Qu.:2.000	1st Qu.:5.000	1st Qu.:0.0000	1st Qu.:1.00
##	Median :3.000	Median :6.000	Median :1.0000	Median :2.00
##	Mean :3.259	Mean :6.183	Mean :0.8835	Mean :2.29
##	3rd Qu.:4.000	3rd Qu.:7.000	3rd Qu.:1.0000	3rd Qu.:3.00
##	Max. :9.000	Max. :9.000	Max. :7.0000	Max. :9.00
##	MFALLEEN	MFGEKIND	MFWEKIND	MOPLHOOG
##	Min. :0.000	Min. :0.00	Min. :0.0	Min. :0.000
##	1st Qu.:0.000	1st Qu.:2.00	1st Qu.:3.0	1st Qu.:0.000
##	Median :2.000	Median :3.00	Median :4.0	Median :1.000
##	Mean :1.888	Mean :3.23	Mean :4.3	Mean :1.461
##	3rd Qu.:3.000	3rd Qu.:4.00	3rd Qu.:6.0	3rd Qu.:2.000
##	Max. :9.000	Max. :9.00	Max. :9.0	Max. :9.000
##	MOPLMIDD	MOPLLAAG	MBERHOOG	MBERZELF
##	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
##	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:0.000	1st Qu.:0.000
##	Median :3.000	Median :5.000	Median :2.000	Median :0.000
##	Mean :3.351	Mean :4.572	Mean :1.895	Mean :0.398
##	3rd Qu.:4.000	3rd Qu.:6.000	3rd Qu.:3.000	3rd Qu.:1.000
##	Max. :9.000	Max. :9.000	Max. :9.000	Max. :5.000
##	MBERBOER	MBERMIDD	MBERARBG	MBERARBO
##	Min. :0.0000	Min. :0.000	Min. :0.00	Min. :0.000
##	1st Qu.:0.0000	1st Qu.:2.000	1st Qu.:1.00	1st Qu.:1.000
##	Median :0.0000	Median :3.000	Median :2.00	Median :2.000
##	Mean :0.5223	Mean :2.899	Mean :2.22	Mean :2.306
##	3rd Qu.:1.0000	3rd Qu.:4.000	3rd Qu.:3.00	3rd Qu.:3.000

##	Max. :9.0000	Max. :9.000	Max. :9.00	Max. :9.000
##	MSKA	MSKB1	MSKB2	MSKC
##	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
##	1st Qu.:0.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
##	Median :1.000	Median :2.000	Median :2.000	Median :4.000
##	Mean :1.621	Mean :1.607	Mean :2.203	Mean :3.759
##	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:5.000
##	Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.000
##	MSKD	MHHUUR	MHKOOP	MAUT1
##	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.00
##	1st Qu.:0.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:5.00
##	Median :1.000	Median :4.000	Median :5.000	Median :6.00
##	Mean :1.067	Mean :4.237	Mean :4.772	Mean :6.04
##	3rd Qu.:2.000	3rd Qu.:7.000	3rd Qu.:7.000	3rd Qu.:7.00
##	Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.00
##	MAUT2	MAUTO	MZFONDS	MZPART
##	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.0000
##	1st Qu.:0.000	1st Qu.:1.000	1st Qu.:5.000	1st Qu.:1.000
##	Median :1.000	Median :2.000	Median :7.000	Median :2.000
##	Mean :1.316	Mean :1.959	Mean :6.277	Mean :2.729
##	3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:8.000	3rd Qu.:4.000
##	Max. :7.000	Max. :9.000	Max. :9.000	Max. :9.000
##	MINKM30	MINK3045	MINK4575	MINK7512
##	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.0000
##	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:0.0000
##	Median :2.000	Median :4.000	Median :3.000	Median :0.0000
##	Mean :2.574	Mean :3.536	Mean :2.731	Mean :0.7961
##	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:1.0000
##	Max. :9.000	Max. :9.000	Max. :9.000	Max. :9.0000
##	MINK123M	MINKGEM	MKOOPKLA	PWAPART
##	Min. :0.0000	Min. :0.000	Min. :1.000	Min. :0.0000
##	1st Qu.:0.0000	1st Qu.:3.000	1st Qu.:3.000	1st Qu.:0.0000
##	Median :0.0000	Median :4.000	Median :4.000	Median :0.0000
##	Mean :0.2027	Mean :3.784	Mean :4.236	Mean :0.7712
##	3rd Qu.:0.0000	3rd Qu.:4.000	3rd Qu.:6.000	3rd Qu.:2.0000
##	Max. :9.0000	Max. :9.000	Max. :8.000	Max. :3.0000
##	PWABEDR	PWALAND	PPERSAUT	PBESAUT
##	Min. :0.00000	Min. :0.00000	Min. :0.00	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00	1st Qu.:0.00000
##	Median :0.00000	Median :0.00000	Median :5.00	Median :0.00000
##	Mean :0.04002	Mean :0.07162	Mean :2.97	Mean :0.04827
##	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:6.00	3rd Qu.:0.00000
##	Max. :6.00000	Max. :4.00000	Max. :8.00	Max. :7.00000
##	PMOTSCO	PVRAAUT	PAANHANG	PTRACTOR
##	Min. :0.0000	Min. :0.000000	Min. :0.00000	Min. :0.00000
##	1st Qu.:0.0000	1st Qu.:0.000000	1st Qu.:0.00000	1st Qu.:0.00000
##	Median :0.0000	Median :0.000000	Median :0.00000	Median :0.00000
##	Mean :0.1754	Mean :0.009447	Mean :0.02096	Mean :0.09258
##	3rd Qu.:0.0000	3rd Qu.:0.000000	3rd Qu.:0.00000	3rd Qu.:0.00000
##	Max. :7.0000	Max. :9.000000	Max. :5.00000	Max. :6.00000
##	PWERKT	PBROM	PLEVEN	PPERSONG
##	Min. :0.00000	Min. :0.000	Min. :0.0000	Min. :0.00000
##	1st Qu.:0.00000	1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:0.00000
##	Median :0.00000	Median :0.000	Median :0.0000	Median :0.00000

##	Mean	:0.01305	Mean	:0.215	Mean	:0.1948	Mean	:0.01374
##	3rd Qu.:	0.00000	3rd Qu.:	0.000	3rd Qu.:	0.0000	3rd Qu.:	0.00000
##	Max.	:6.00000	Max.	:6.000	Max.	:9.0000	Max.	:6.00000
##	PGEZONG		PWAOREG		PBRAND		PZEILPL	
##	Min.	:0.00000	Min.	:0.00000	Min.	:0.000	Min.	:0.0000000
##	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.000	1st Qu.:	0.0000000
##	Median	:0.00000	Median	:0.00000	Median	:2.000	Median	:0.0000000
##	Mean	:0.01529	Mean	:0.02353	Mean	:1.828	Mean	:0.0008588
##	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	4.000	3rd Qu.:	0.0000000
##	Max.	:3.00000	Max.	:7.00000	Max.	:8.000	Max.	:3.0000000
##	PPLEZIER		PFIETS		PINBOED		PBYSTAND	
##	Min.	:0.00000	Min.	:0.00000	Min.	:0.00000	Min.	:0.00000
##	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.00000
##	Median	:0.00000	Median	:0.00000	Median	:0.00000	Median	:0.00000
##	Mean	:0.01889	Mean	:0.02525	Mean	:0.01563	Mean	:0.04758
##	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	0.00000
##	Max.	:6.00000	Max.	:1.00000	Max.	:6.00000	Max.	:5.00000
##	AWAPART		AWABEDR		AWALAND		APERSAUT	
##	Min.	:0.000	Min.	:0.00000	Min.	:0.00000	Min.	:0.0000
##	1st Qu.:	0.000	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.0000
##	Median	:0.000	Median	:0.00000	Median	:0.00000	Median	:1.0000
##	Mean	:0.403	Mean	:0.01477	Mean	:0.02061	Mean	:0.5622
##	3rd Qu.:	1.000	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	1.0000
##	Max.	:2.000	Max.	:5.00000	Max.	:1.00000	Max.	:7.0000
##	ABESAUT		AMOTSCO		AVRAAUT		AAANHANG	
##	Min.	:0.00000	Min.	:0.00000	Min.	:0.000000	Min.	:0.00000
##	1st Qu.:	0.00000	1st Qu.:	0.00000	1st Qu.:	0.000000	1st Qu.:	0.00000
##	Median	:0.00000	Median	:0.00000	Median	:0.000000	Median	:0.00000
##	Mean	:0.01048	Mean	:0.04105	Mean	:0.002233	Mean	:0.01254
##	3rd Qu.:	0.00000	3rd Qu.:	0.00000	3rd Qu.:	0.000000	3rd Qu.:	0.00000
##	Max.	:4.00000	Max.	:8.00000	Max.	:3.000000	Max.	:3.00000
##	ATTRACTOR		AWERKT		ABROM		ALEVEN	
##	Min.	:0.00000	Min.	:0.000000	Min.	:0.00000	Min.	:0.00000
##	1st Qu.:	0.00000	1st Qu.:	0.000000	1st Qu.:	0.00000	1st Qu.:	0.00000
##	Median	:0.00000	Median	:0.000000	Median	:0.00000	Median	:0.00000
##	Mean	:0.03367	Mean	:0.006183	Mean	:0.07042	Mean	:0.07661
##	3rd Qu.:	0.00000	3rd Qu.:	0.000000	3rd Qu.:	0.00000	3rd Qu.:	0.00000
##	Max.	:4.00000	Max.	:6.000000	Max.	:2.00000	Max.	:8.00000
##	APERSONG		AGEZONG		AWAOREG		ABRAND	
##	Min.	:0.000000	Min.	:0.000000	Min.	:0.000000	Min.	:0.0000
##	1st Qu.:	0.000000	1st Qu.:	0.000000	1st Qu.:	0.000000	1st Qu.:	0.0000
##	Median	:0.000000	Median	:0.000000	Median	:0.000000	Median	:1.0000
##	Mean	:0.005325	Mean	:0.006527	Mean	:0.004638	Mean	:0.5701
##	3rd Qu.:	0.000000	3rd Qu.:	0.000000	3rd Qu.:	0.000000	3rd Qu.:	1.0000
##	Max.	:1.000000	Max.	:1.000000	Max.	:2.000000	Max.	:7.0000
##	AZEILPL		APLEZIER		AFIETS			
##	Min.	:0.0000000	Min.	:0.000000	Min.	:0.00000		
##	1st Qu.:	0.0000000	1st Qu.:	0.000000	1st Qu.:	0.00000		
##	Median	:0.0000000	Median	:0.000000	Median	:0.00000		
##	Mean	:0.0005153	Mean	:0.006012	Mean	:0.03178		
##	3rd Qu.:	0.0000000	3rd Qu.:	0.000000	3rd Qu.:	0.00000		
##	Max.	:1.0000000	Max.	:2.000000	Max.	:3.00000		
##	AINBOED		ABYSTAND		Purchase			
##	Min.	:0.000000	Min.	:0.00000	No	:5474		

```
## 1st Qu.:0.000000 1st Qu.:0.000000 Yes: 348
## Median :0.000000 Median :0.000000
## Mean :0.007901 Mean :0.01426
## 3rd Qu.:0.000000 3rd Qu.:0.000000
## Max. :2.000000 Max. :2.000000
```

```
plot(Caravan$Purchase)
```



Data engineering :

In this stage the NA values of the data set has been detected and the rows with such values are omitted from the data set.

```
## 'data.frame': 5822 obs. of 86 variables:
## $ MOSTYPE : num 33 37 37 9 40 23 39 33 33 11 ...
## $ MAANTHUI: num 1 1 1 1 1 1 2 1 1 2 ...
## $ MGEMOMV : num 3 2 2 3 4 2 3 2 2 3 ...
## $ MGEMLEEF: num 2 2 2 3 2 1 2 3 4 3 ...
## $ MOSHOOFD: num 8 8 8 3 10 5 9 8 8 3 ...
## $ MGODRK : num 0 1 0 2 1 0 2 0 0 3 ...
## $ MGODPR : num 5 4 4 3 4 5 2 7 1 5 ...
## $ MGODOV : num 1 1 2 2 1 0 0 0 3 0 ...
## $ MGODGE : num 3 4 4 4 4 5 5 2 6 2 ...
## $ MRELGE : num 7 6 3 5 7 0 7 7 6 7 ...
## $ MRELSA : num 0 2 2 2 1 6 2 2 0 0 ...
## $ MRELOV : num 2 2 4 2 2 3 0 0 3 2 ...
## $ MFALLEEN: num 1 0 4 2 2 3 0 0 3 2 ...
```

```

## $ MFGEKIND: num 2 4 4 3 4 5 3 5 3 2 ...
## $ MFW EKIND: num 6 5 2 4 4 2 6 4 3 6 ...
## $ MOPLHOOG: num 1 0 0 3 5 0 0 0 0 0 ...
## $ MOPLMIDD: num 2 5 5 4 4 5 4 3 1 4 ...
## $ MOPLLAAG: num 7 4 4 2 0 4 5 6 8 5 ...
## $ MBERHOOG: num 1 0 0 4 0 2 0 2 1 2 ...
## $ MBERZELF: num 0 0 0 0 5 0 0 0 1 0 ...
## $ MBERBOER: num 1 0 0 0 4 0 0 0 0 0 ...
## $ MBERMIDD: num 2 5 7 3 0 4 4 2 1 3 ...
## $ MBERARBG: num 5 0 0 1 0 2 1 5 8 3 ...
## $ MBERARBO: num 2 4 2 2 0 2 5 2 1 3 ...
## $ MSKA      : num 1 0 0 3 9 2 0 2 1 1 ...
## $ MSKB1     : num 1 2 5 2 0 2 1 1 1 2 ...
## $ MSKB2     : num 2 3 0 1 0 2 4 2 0 1 ...
## $ MSKC      : num 6 5 4 4 0 4 5 5 8 4 ...
## $ MSKD      : num 1 0 0 0 0 2 0 2 1 2 ...
## $ MHHUUR    : num 1 2 7 5 4 9 6 0 9 0 ...
## $ MHKOOP    : num 8 7 2 4 5 0 3 9 0 9 ...
## $ MAUT1     : num 8 7 7 9 6 5 8 4 5 6 ...
## $ MAUT2     : num 0 1 0 0 2 3 0 4 2 1 ...
## $ MAUTO     : num 1 2 2 0 1 3 1 2 3 2 ...
## $ MZFONDS   : num 8 6 9 7 5 9 9 6 7 6 ...
## $ MZPART    : num 1 3 0 2 4 0 0 3 2 3 ...
## $ MINKM30    : num 0 2 4 1 0 5 4 2 7 2 ...
## $ MINK3045: num 4 0 5 5 0 2 3 5 2 3 ...
## $ MINK4575: num 5 5 0 3 9 3 3 3 1 3 ...
## $ MINK7512: num 0 2 0 0 0 0 0 0 0 1 ...
## $ MINK123M: num 0 0 0 0 0 0 0 0 0 0 ...
## $ MINKGEM   : num 4 5 3 4 6 3 3 3 2 4 ...
## $ MKOOPKLA: num 3 4 4 4 3 3 5 3 3 7 ...
## $ PWAPART   : num 0 2 2 0 0 0 0 0 0 2 ...
## $ PWABEDR   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PWALAND   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PPERSAUT: num 6 0 6 6 0 6 6 0 5 0 ...
## $ PBESAUT   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PMOTSCO   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PVRAAUT   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PAANHANG: num 0 0 0 0 0 0 0 0 0 0 ...
## $ PTRACTOR: num 0 0 0 0 0 0 0 0 0 0 ...
## $ PWERKT    : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PBROM     : num 0 0 0 0 0 0 0 3 0 0 ...
## $ PLEVEN    : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PPERSONG: num 0 0 0 0 0 0 0 0 0 0 ...
## $ PGEZONG   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PWAOREG   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PBRAND    : num 5 2 2 2 6 0 0 0 0 3 ...
## $ PZEILPL   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PPLEZIER: num 0 0 0 0 0 0 0 0 0 0 ...
## $ PFIETS    : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PINBOED   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ PBYSTAND: num 0 0 0 0 0 0 0 0 0 0 ...
## $ AWAPART   : num 0 2 1 0 0 0 0 0 0 1 ...
## $ AWABEDR   : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AWALAND   : num 0 0 0 0 0 0 0 0 0 0 ...

```

```
## $ APERSAUT: num 1 0 1 1 0 1 1 0 1 0 ...
## $ ABESAUT : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AMOTSCO : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AVRAAUT : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AAANHANG: num 0 0 0 0 0 0 0 0 0 0 ...
## $ ATRACTOR: num 0 0 0 0 0 0 0 0 0 0 ...
## $ AWERKT : num 0 0 0 0 0 0 0 0 0 0 ...
## $ ABROM : num 0 0 0 0 0 0 0 1 0 0 ...
## $ ALEVEN : num 0 0 0 0 0 0 0 0 0 0 ...
## $ APERSONG: num 0 0 0 0 0 0 0 0 0 0 ...
## $ AGEZONG : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AWAOREG : num 0 0 0 0 0 0 0 0 0 0 ...
## $ ABRAND : num 1 1 1 1 1 0 0 0 0 1 ...
## $ AZEILPL : num 0 0 0 0 0 0 0 0 0 0 ...
## $ APLEZIER: num 0 0 0 0 0 0 0 0 0 0 ...
## $ AFIETS : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AINBOED : num 0 0 0 0 0 0 0 0 0 0 ...
## $ ABYSTAND: num 0 0 0 0 0 0 0 0 0 0 ...
## $ Purchase: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
```

#Train/Test Split:

Use caret for training machine learning models

##Use GLM to fit the model

```
#caret default
caretControl <- trainControl(method = "boot",number = 25, classProbs = TRUE)
```

```
set.seed(3233)
modelGLM <- train(as.factor(Purchase)~., data = train, method = "glm",
                  trControl=caretControl)
summary(modelGLM)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9424  -0.3627  -0.2388  -0.1529   3.2856
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.599e+02  1.199e+04   0.022  0.98271
## MOSTYPE      1.308e-01  5.683e-02   2.301  0.02139 *
## MAANTHUI     -2.718e-02  2.062e-01  -0.132  0.89514
## MGEMOMV      -6.232e-02  1.681e-01  -0.371  0.71080
## MGEMLEEF      2.941e-01  1.183e-01   2.487  0.01288 *
## MOSHOOFD     -5.448e-01  2.542e-01  -2.144  0.03207 *
## MGODRK       -6.345e-02  1.275e-01  -0.498  0.61881
## MGODPR       -3.275e-03  1.390e-01  -0.024  0.98120
## MGODOV        8.202e-03  1.249e-01   0.066  0.94765
## MGODGE       -3.805e-02  1.324e-01  -0.287  0.77388
## MRELGE        1.999e-01  1.909e-01   1.047  0.29520
## MRELSA        3.852e-02  1.794e-01   0.215  0.83000
```

## MRELOV	5.647e-02	1.906e-01	0.296	0.76701
## MFALLEEN	-9.595e-02	1.590e-01	-0.603	0.54621
## MFGEKIND	-1.679e-01	1.621e-01	-1.036	0.30020
## MFW EKIND	-1.177e-01	1.718e-01	-0.685	0.49319
## MOPLHOOG	-1.615e-01	1.602e-01	-1.008	0.31356
## MOPLMIDD	-2.573e-01	1.650e-01	-1.559	0.11896
## MOPLLAAG	-4.087e-01	1.658e-01	-2.465	0.01369 *
## MBERHOOG	1.129e-01	1.117e-01	1.010	0.31250
## MBERZELF	6.836e-02	1.174e-01	0.582	0.56043
## MBERBOER	-9.021e-02	1.303e-01	-0.692	0.48885
## MBERMIDD	9.068e-02	1.087e-01	0.834	0.40436
## MBERARBG	3.772e-02	1.070e-01	0.352	0.72455
## MBERARBO	8.288e-02	1.090e-01	0.761	0.44684
## MSKA	-1.863e-02	1.219e-01	-0.153	0.87852
## MSKB1	-2.612e-02	1.201e-01	-0.217	0.82788
## MSKB2	3.277e-02	1.065e-01	0.308	0.75844
## MSKC	1.361e-01	1.173e-01	1.161	0.24579
## MSKD	3.985e-03	1.156e-01	0.034	0.97250
## MHHUUR	-1.474e+01	9.021e+02	-0.016	0.98697
## MHKOOP	-1.473e+01	9.021e+02	-0.016	0.98697
## MAUT1	-1.230e-01	1.881e-01	-0.654	0.51312
## MAUT2	-1.405e-01	1.728e-01	-0.813	0.41615
## MAUTO	-1.784e-01	1.805e-01	-0.989	0.32283
## MZFONDS	-1.450e+01	9.808e+02	-0.015	0.98820
## MZPART	-1.454e+01	9.808e+02	-0.015	0.98817
## MINKM30	4.941e-02	1.198e-01	0.412	0.68009
## MINK3045	8.857e-02	1.148e-01	0.772	0.44033
## MINK4575	6.372e-02	1.157e-01	0.551	0.58178
## MINK7512	6.637e-02	1.233e-01	0.538	0.59050
## MINK123M	-2.482e-01	1.788e-01	-1.388	0.16506
## MINKGEM	1.013e-01	1.211e-01	0.836	0.40301
## MKOOPKLA	1.086e-01	5.588e-02	1.944	0.05193 .
## PWAPART	7.080e-01	4.654e-01	1.521	0.12820
## PWABEDR	-3.508e-01	8.195e-01	-0.428	0.66859
## PWALAND	-4.794e-01	1.005e+00	-0.477	0.63349
## PPERSAUT	1.662e-01	5.331e-02	3.118	0.00182 **
## PBESAUT	1.215e+01	4.800e+02	0.025	0.97981
## PMOTSCO	-3.438e-01	2.856e-01	-1.204	0.22865
## PVRAAUT	-1.804e+00	2.167e+03	-0.001	0.99934
## PAANHANG	5.018e-01	1.066e+00	0.471	0.63776
## PTRACTOR	4.494e-01	5.023e-01	0.895	0.37101
## PWERKT	-3.929e+00	3.265e+03	-0.001	0.99904
## PBROM	3.254e-02	6.804e-01	0.048	0.96185
## PLEVEN	-2.313e-01	1.312e-01	-1.762	0.07800 .
## PPERSONG	-3.060e-01	1.810e+00	-0.169	0.86572
## PGEZONG	5.831e-01	1.211e+00	0.482	0.63011
## PWAOREG	1.427e+00	1.387e+00	1.029	0.30353
## PBRAND	1.954e-01	9.059e-02	2.157	0.03103 *
## PZEILPL	-7.789e-02	4.612e+03	0.000	0.99999
## PPLEZIER	-1.279e-01	3.958e-01	-0.323	0.74653
## PFIETS	4.646e-01	9.753e-01	0.476	0.63384
## PINBOED	-1.087e+00	9.800e-01	-1.109	0.26726
## PBYSTAND	-4.386e-01	6.103e-01	-0.719	0.47237
## AWAPART	-1.192e+00	9.353e-01	-1.275	0.20239

```

## AWABEDR      7.530e-01  2.253e+00  0.334  0.73816
## AWALAND      8.137e-01  3.453e+00  0.236  0.81368
## APERSAUT      2.988e-01  2.306e-01  1.295  0.19519
## ABESAUT     -7.290e+01  2.880e+03 -0.025  0.97981
## AMOTSCO      1.599e+00  1.142e+00  1.399  0.16171
## AVRAAUT     -5.901e+00  9.497e+03 -0.001  0.99950
## AAANHANG     -3.665e-01  1.850e+00 -0.198  0.84294
## ATRACTOR     -1.464e+00  1.644e+00 -0.891  0.37315
## AWERKT      -4.492e+00  6.871e+03 -0.001  0.99948
## ABROM       -7.082e-01  2.056e+00 -0.344  0.73051
## ALEVEN       6.383e-01  2.560e-01  2.494  0.01264 *
## APERSONG      1.180e-01  3.807e+00  0.031  0.97527
## AGEZONG     -9.409e-01  3.077e+00 -0.306  0.75974
## AWAOREG     -6.292e+00  8.348e+00 -0.754  0.45101
## ABRAND      -2.360e-01  3.222e-01 -0.732  0.46393
## AZEILPL     -1.509e+01  1.031e+04 -0.001  0.99883
## APLEZIER      2.679e+00  1.326e+00  2.021  0.04329 *
## AFIETS       2.699e-01  6.676e-01  0.404  0.68594
## AINBOED      2.497e+00  1.616e+00  1.545  0.12225
## ABYSTAND      2.072e+00  2.182e+00  0.950  0.34217
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1881.6  on 4131  degrees of freedom
## Residual deviance: 1574.0  on 4046  degrees of freedom
## AIC: 1746
##
## Number of Fisher Scoring iterations: 17
modelGLM

## Generalized Linear Model
##
## 4132 samples
## 85 predictor
## 2 classes: 'No', 'Yes'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 4132, 4132, 4132, 4132, 4132, 4132, ...
## Resampling results:
##
## Accuracy   Kappa
## 0.9353473  0.02731244

#turn probabilities into classes and look at their frequencies:
p_modelGLM <- predict(modelGLM, test, type="prob")
p_ClassGLM <- predict(modelGLM, test)
confusionMatrix(p_ClassGLM, factor(test$Purchase))

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes

```



```

##          No  1574   96
##          Yes   17    3
##
##              Accuracy : 0.9331
##              95% CI : (0.9202, 0.9446)
##          No Information Rate : 0.9414
##          P-Value [Acc > NIR] : 0.9312
##
##              Kappa : 0.0313
##  McNemar's Test P-Value : 2.174e-13
##
##          Sensitivity : 0.9893
##          Specificity : 0.0303
##          Pos Pred Value : 0.9425
##          Neg Pred Value : 0.1500
##          Prevalence : 0.9414
##          Detection Rate : 0.9314
##          Detection Prevalence : 0.9882
##          Balanced Accuracy : 0.5098
##
##          'Positive' Class : No
##
##SVM ##SVMLinear
set.seed(3233)
modelSVMLinear <- train(factor(Purchase)~., data = train, method = "svmLinear",
                        trControl=caretControl, Probs = TRUE)
summary(modelSVMLinear)

## Length Class Mode
##      1  ksvm   S4
modelSVMLinear

## Support Vector Machines with Linear Kernel
##
## 4132 samples
## 85 predictor
## 2 classes: 'No', 'Yes'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 4132, 4132, 4132, 4132, 4132, 4132, ...
## Resampling results:
##
## Accuracy Kappa
## 0.9411778 0.006622094
##
## Tuning parameter 'C' was held constant at a value of 1
p_modelSVMLinear <- predict(modelSVMLinear, test, "prob")
p_ClassSVMLinear <- predict(modelSVMLinear, test)
confusionMatrix(p_ClassSVMLinear, factor(test$Purchase))

## Confusion Matrix and Statistics
##

```

```

##           Reference
## Prediction   No  Yes
##           No 1591  99
##           Yes   0   0
##
##           Accuracy : 0.9414
##           95% CI : (0.9291, 0.9521)
##           No Information Rate : 0.9414
##           P-Value [Acc > NIR] : 0.5267
##
##           Kappa : 0
## Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 1.0000
##           Specificity : 0.0000
##           Pos Pred Value : 0.9414
##           Neg Pred Value :      NaN
##           Prevalence : 0.9414
##           Detection Rate : 0.9414
##           Detection Prevalence : 1.0000
##           Balanced Accuracy : 0.5000
##
##           'Positive' Class : No
##
set.seed(3233)
modelSVMRadial <- train(factor(Purchase)~., data = train, method = "svmRadial",
                        trControl=caretControl)
summary(modelSVMRadial)

## Length Class Mode
##      1  ksvm   S4

modelSVMRadial

## Support Vector Machines with Radial Basis Function Kernel
##
## 4132 samples
## 85 predictor
## 2 classes: 'No', 'Yes'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 4132, 4132, 4132, 4132, 4132, 4132, ...
## Resampling results across tuning parameters:
##
##  C      Accuracy  Kappa
##  0.25  0.9335167  0.04148473
##  0.50  0.9332993  0.03902150
##  1.00  0.9342537  0.03932611
##
## Tuning parameter 'sigma' was held constant at a value of 0.01027934
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were sigma = 0.01027934 and C = 1.

```

```

p_modelSVMRadial <- predict(modelSVMRadial, test, "prob")
p_ClasslSVMRadial <- predict(modelSVMRadial, test)
confusionMatrix(p_ClasslSVMRadial, factor(test$Purchase))

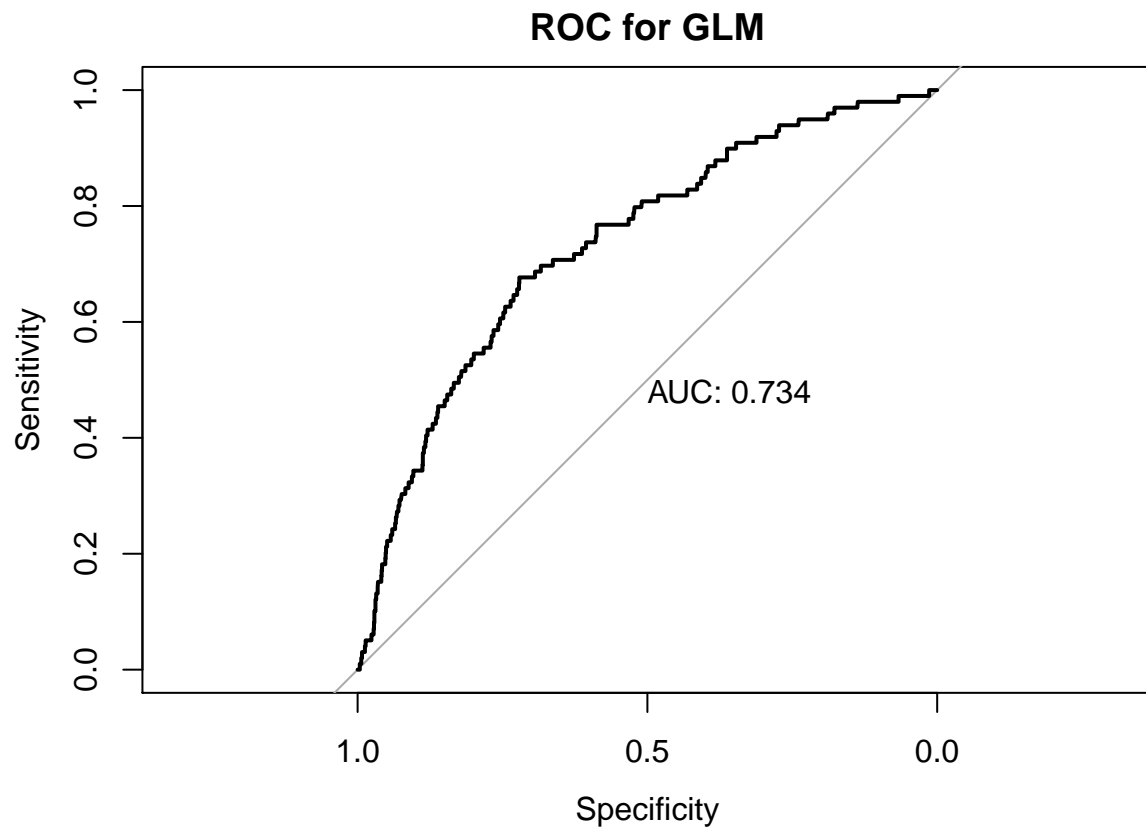
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   No  Yes
##           No 1587  97
##           Yes   4   2
##
##           Accuracy : 0.9402
##           95% CI : (0.9279, 0.9511)
##           No Information Rate : 0.9414
##           P-Value [Acc > NIR] : 0.6076
##
##           Kappa : 0.0316
##           Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.9975
##           Specificity : 0.0202
##           Pos Pred Value : 0.9424
##           Neg Pred Value : 0.3333
##           Prevalence : 0.9414
##           Detection Rate : 0.9391
##           Detection Prevalence : 0.9964
##           Balanced Accuracy : 0.5088
##
##           'Positive' Class : No
##

comparison<-resamples(list(GLM=modelGLM, SVMLinear=modelSVMLinear, SVMRadial=modelSVMRadial))
summary(comparison)

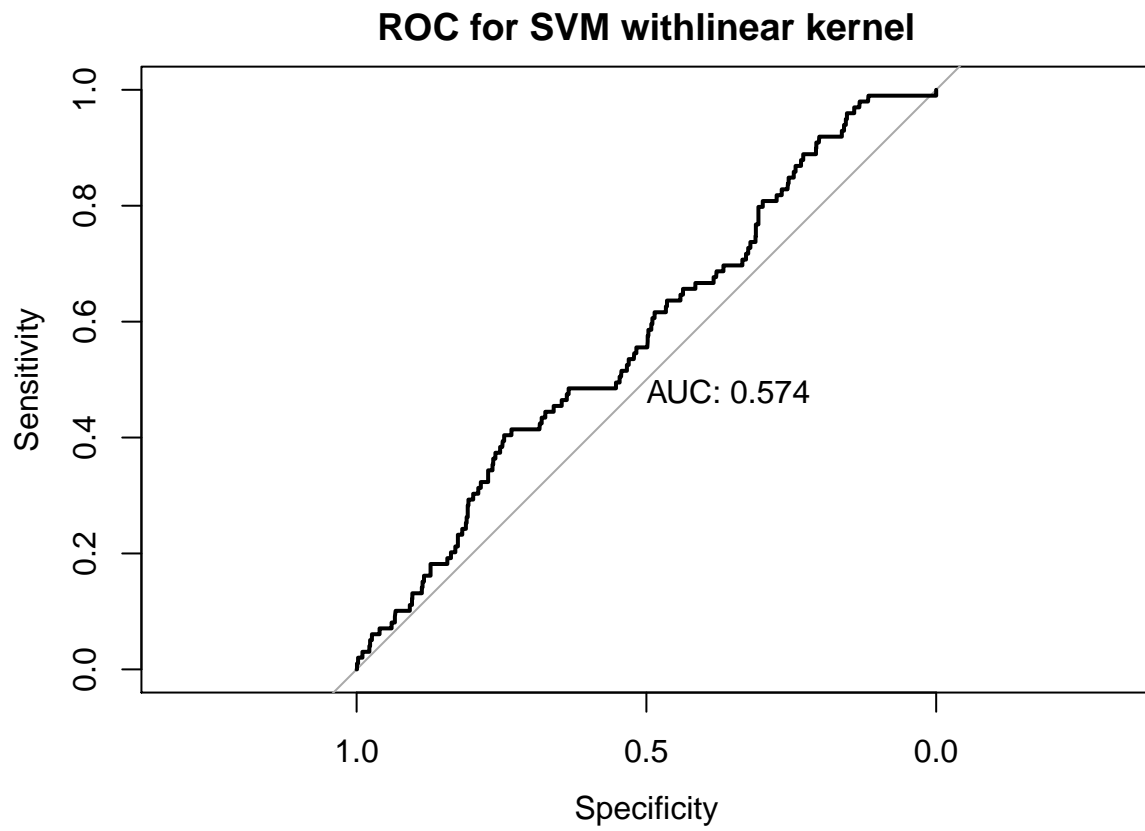
##
## Call:
## summary.resamples(object = comparison)
##
## Models: GLM, SVMLinear, SVMRadial
## Number of resamples: 25
##
## Accuracy
##           Min.    1st Qu.    Median    Mean    3rd Qu.    Max. NA's
## GLM           0.9202373 0.9332452 0.9361277 0.9353473 0.9379636 0.9456891    0
## SVMLinear     0.9340804 0.9381720 0.9405155 0.9411778 0.9427811 0.9540682    0
## SVMRadial     0.9254032 0.9308996 0.9338235 0.9342537 0.9373351 0.9422946    0
##
## Kappa
##           Min.    1st Qu.    Median    Mean    3rd Qu.
## GLM          -0.016291424 0.01144304 0.02367364 0.027312443 0.04252269
## SVMLinear    -0.001306781 0.00000000 0.00000000 0.006622094 0.00000000
## SVMRadial     0.001845787 0.02501580 0.03851417 0.039326108 0.05393764
##           Max. NA's
## GLM           0.10681361    0

```

```
## SVMLinear 0.07056305 0
## SVMRadial 0.08820103 0
ROC_GLM <- roc(factor(test$Purchase)~p_modelGLM[,2], plot = TRUE, print.auc = TRUE, main="ROC for GLM")
```



```
ROC_SVMlinear<- roc(factor(test$Purchase)~p_modelSVMLinear[,2], plot = TRUE, print.auc = TRUE, main="ROC for SVMLinear")
```



```
ROC_SVMRadial<- roc(factor(test$Purchase)~p_modelSVMRadial[,2], plot = TRUE, print.auc = TRUE, main="ROC for SVM with linear kernel")
```

