Youden’s Index, Lift and Gain

German Credit Data

Lecture(Practical 17)

Youden's Index Calculation

- To calculate Youden's index, one has to first calculate the value of sensitivity(p)+specificity(p)-1 for different classification cut off probability p.

Classification Cut-off probability	Sensitivity(p)	Specificity(p)	sensitivity(p)+specificity(p)-1
0.05			
0.1			
…			
…			
0.9			
0.95			

- The maximum value of sensitivity(p)+specificity(p)-1 is chosen and the corresponding classification cut-off probability may be considered as optimal classification cut-off probability using Youden's Index .

Youden’s Index (with cut-off 0.5)#german credit data

- youden(tp, fp, tn, fn, ...)

```
          actual
predicted   0   1
        0 541 207
        1  20  32
```

- youden(32, 20, 541, 207)
- #youden_index = sensitivity + specificity - 1
- youden_index = 0.134+0.964 - 1
- youden_index
- [1] 0.098
- This process will be repeated for different cut-off to obtain optimal cut-off probability.

Gain and Lift

The gain chart and lift charts are obtained using the following steps.

1. Predict the probability Y=1(positive) using the logistic regression model (LR)and arrange the observation in the decreasing order of predicted probability.[i.e., P(Y=1)].
2. Divide the datasets into deciles . Calculate the number of positives (Y=1) in each decile and cumulative number of positives up to a decile.
3. Gain is the ratio between cumulative number of positive observations up to a decile to total number of positive observations in the data. Gain chart is a chart between gain on the vertical axis and the decile on the horizontal axis.

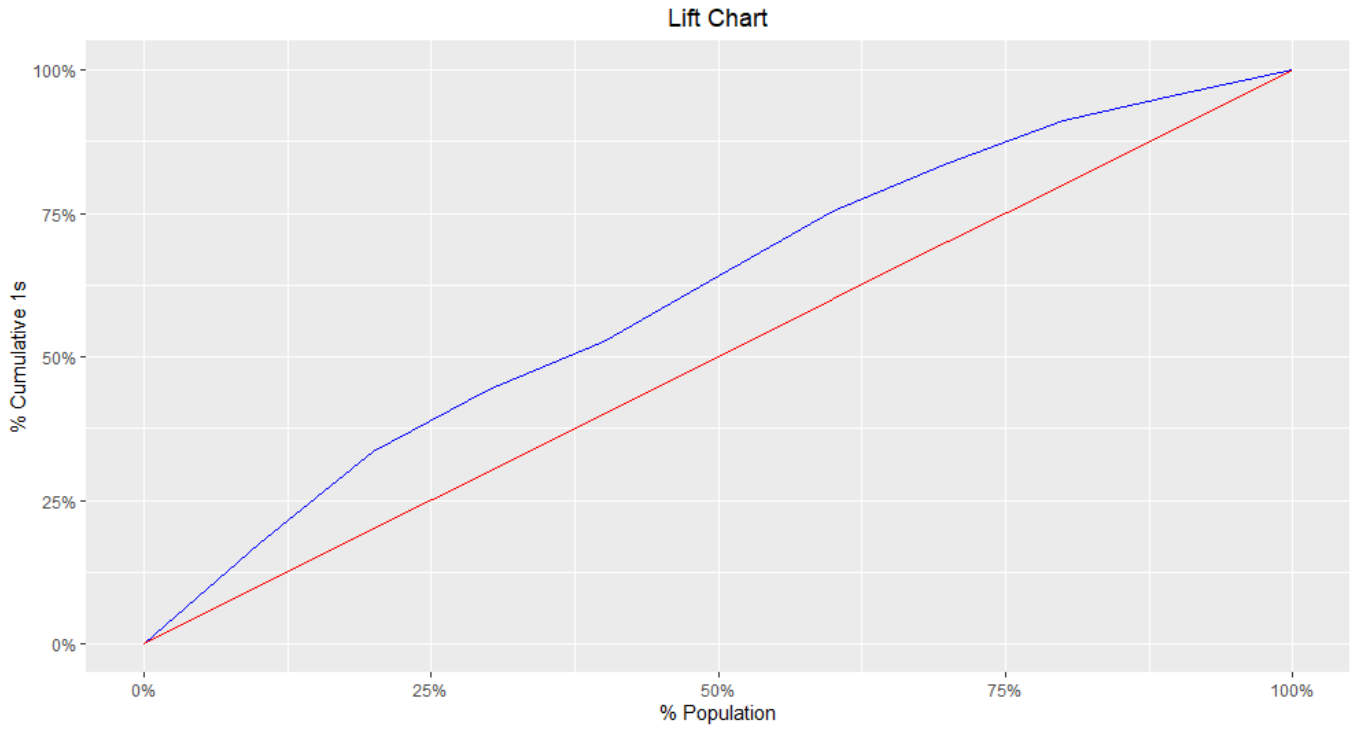
$$\text{Gain}=\frac{Cumulative\ number\ of\ positive\ observations\ upto\ decile\ i}{Total\ number\ of\ positive\ observations\ in\ the\ data}$$

4. Lift is the ratio of number of positive observations up to decile i using the LR model to the expected number of positives up to that decile i based on a random model (not using a model). Lift chart is the chart between lift on vertical axis and the corresponding decile on the horizontal axis.

$$\text{Lift}=\frac{Cumulative\ number\ of\ positive\ observations\ upto\ decile\ i\ using\ LR\ model}{Cumulative\ number\ of\ positive\ observations\ up\ to\ decile\ i\ based\ on\ a\ random\ model}$$

GainChart

- Gain can be interpreted as the gain in identifying customers who are likely to subscribe compared to a random model.
- library(blorr)
- k<-blr_gains_table(lr5)
- #blr_confusion_matrix(lr5)
- plot(k)



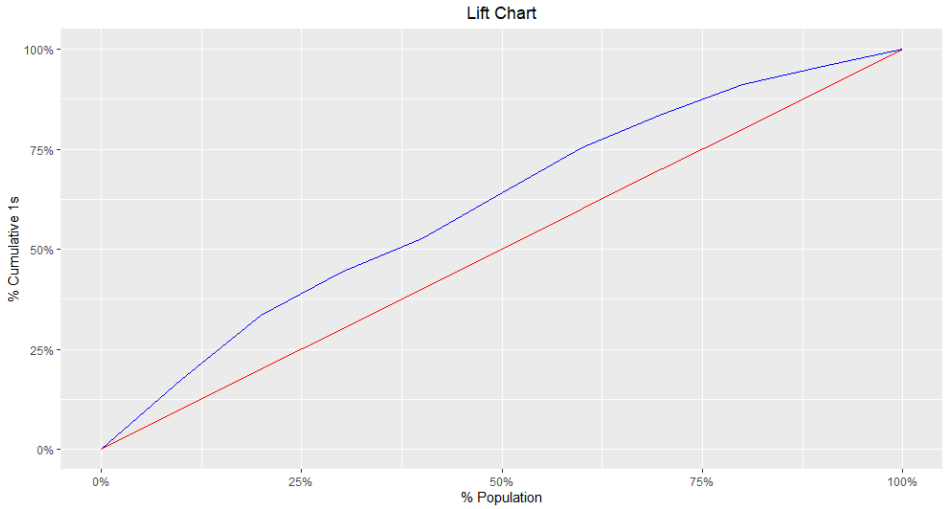
Gain Table

- gt <- blr_gains_table(lr5)

	decile	total	1	0	ks	tp	tn	fp	fn	sensitivity
1	1	80	42	38	10.799603	42	523	38	197	17.57322
2	2	80	38	42	19.212554	80	481	80	159	33.47280
3	3	80	26	54	20.465546	106	427	134	133	44.35146
4	4	80	20	60	18.138560	126	367	194	113	52.71967
5	5	80	27	53	19.988216	153	314	247	86	64.01674
6	6	80	27	53	21.837872	180	261	300	59	75.31381
7	7	80	20	60	19.510885	200	201	360	39	83.68201
8	8	80	18	62	15.990573	218	139	422	21	91.21339
9	9	80	11	69	8.293618	229	70	491	10	95.81590
10	10	80	10	70	0.000000	239	0	561	0	100.00000

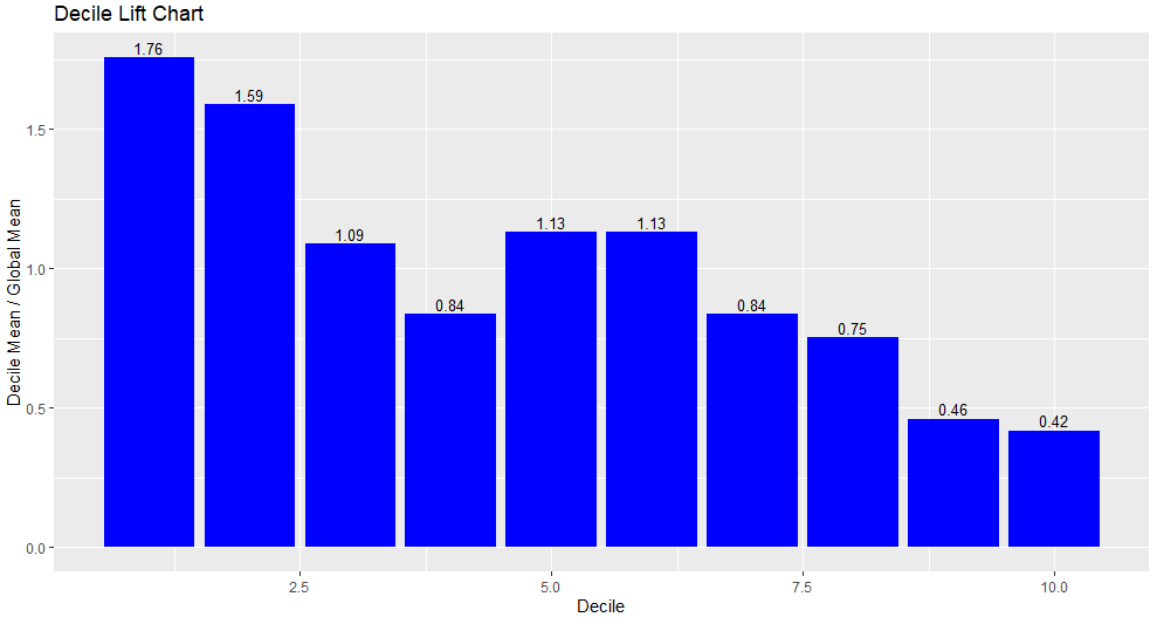
	specificity	accuracy
1	93.22638	70.625
2	85.73975	70.125
3	76.11408	66.625
4	65.41889	61.625
5	55.97148	58.375
6	46.52406	55.125
7	35.82888	50.125
8	24.77718	44.625
9	12.47772	37.375
10	0.00000	29.875

Gain Calculation		
Decile	Gain	Gain %
1	42/239=0.1757	17.57
2	(42+38)/239=0.3347	33.47
…	…	…
5	153/239=0.6401	64.01
…	…	



Lift Chart

- blr_decile_lift_chart(k)



Lift Table	
Decile	Lift
1	1.76
2	1.59
…	
5	1.13
…	

- The response is 17.6% using LR model against 10% in a random model. Thus the lift is 1.76 (17.6/10). That is targeting the customers using the model can capture 1.76 times the number of subscribers compared to a random model in decile 1