K2 Analytics
Building Skills, Building Individuals

Model Performance Measures

- Rajesh Jakhotia

19-Sep-2017

Earning is in Learning
- Rajesh Jakhotia

List of Model Performance Measures

- Rank Ordering
- KS
- Lift Chart
- Concordance – Discordance
- Gini Coefficient
- Classification Error

K2 Analytics
Building Skills, Building Individuals

Rank Ordering

Earning is in Learning
- Rajesh Jakhotia

Rank Ordering Context

- Rank Ordering is often used as one of the key Model Performance Measure in Predictive Modeling
- Predictive Models e.g.
  - In Marketing, predictive models are used to predict the likelihood of customer responding to the marketing offer
  - In Risk, predictive model are used to predict the probability of default of the customer applying for loan
- Rank Ordering as model performance measures helps:
  - assess ability of the model to relatively rank the customers
  - see how well the model separates the Responder Class from the Non-Responder, or the Defaulters from Non-Defaulters and likewise Attriters from Non-Attriters
  - on an ongoing basis track the utility of the model

How does the data look like?

ID	PV1	PV2	PV3	…	…	…	PVn	Target	Probability
1								0	0.0562
2								0	0.1345
.								0	…
.								1	0.5122
.								0	…
								0	…
n								1	…

ID	Unique Identifier for each row, e.g. Customer ID
PV	Predictor / Independent Variables
Target	Flag indicating whether the event occurred or not. E.g. - In case of marketing the 1's indicate customer responded to the offer and 0's indicate the customer did not respond to the offer
Probability	The probability score as computed based on the model. Let us assume that the probabilities given above are for customer responding to the marketing offer

Rank Order Computation Steps | Sorting

Target	Probability
0	0.0562
0	0.1345
0	…
1	0.5122
0	…
0	…
1	…
0	0.0232
0	…
…	…
…	…
…	…

Fields required to create the Rank Order Table

Sort data based on Probability

Target	Probability
0	0.0232
0	…
…	…
1	
…	…
0	0.0562
1	…
0	0.1345
…	…
1	0.5122
…	…
…	…

Low Prob.

High Prob.

Rank Order Computation Steps | Deciling

Target	Probability	Decile
0	0.0232	1
0	…	1
…	…	…
1	…	…
…	…	2
0	0.0562	2
1	…	..
0	0.1345	…
…	…	…
1	0.5122	…
…	…	…
…	…	10

- Deciling is process of splitting the data into 10 buckets each having 10% of the observations

Rank Order Computation Steps | Aggregations

Decile	Base_Cnt	Resp_Cnt
10	1,000	295
9	1,000	176
8	1,000	115
7	1,000	75
6	1,000	35
5	1,000	30
4	1,000	23
3	1,000	18
2	1,000	13
1	1,000	6
Total	10,000	787

- After Deciling, we aggregate the data based on Decile column
- Note: The table here is with some hypothetical numbers to explain the further calculations in Rank Ordering

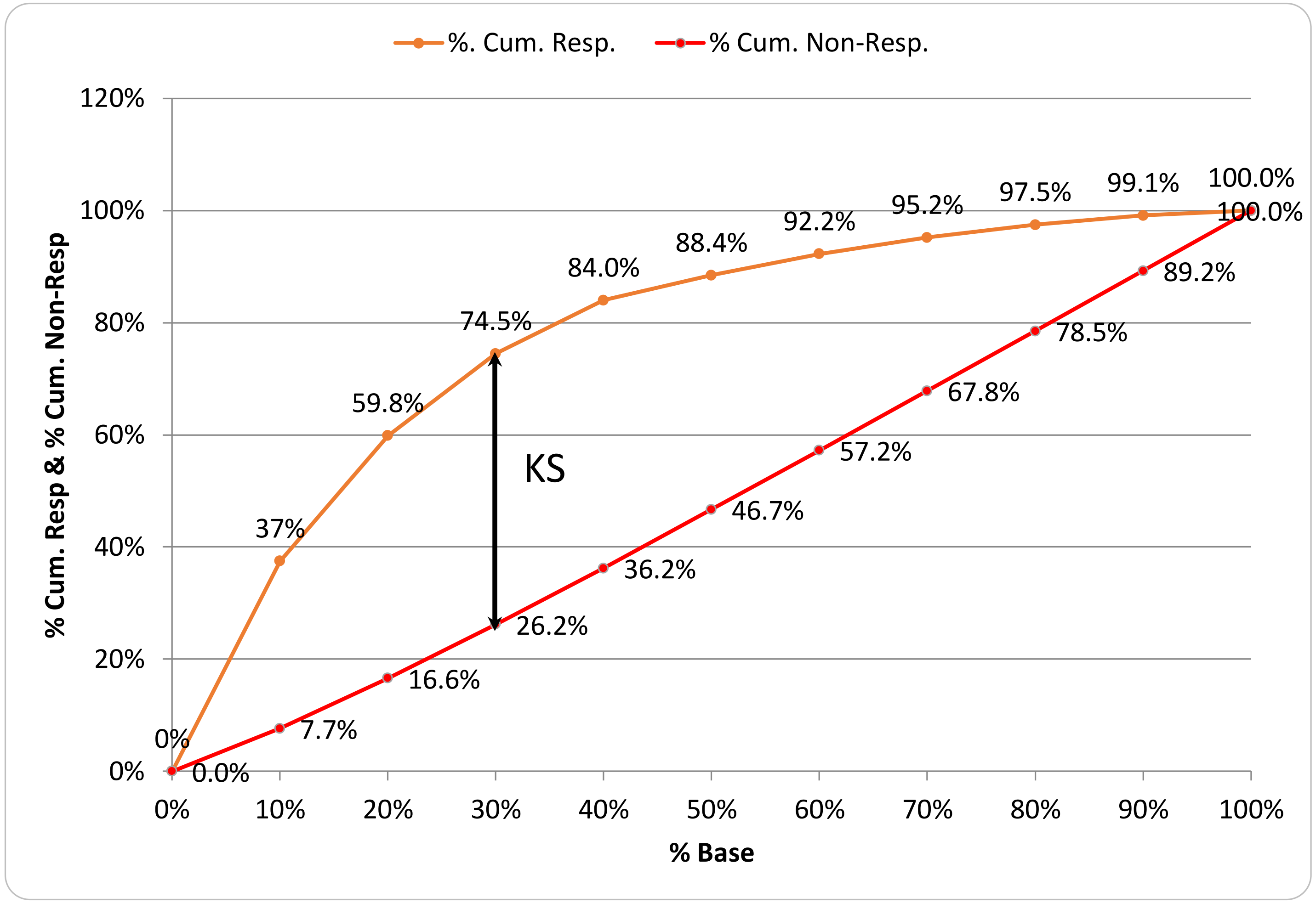
Base_Cnt	Count of Observations in each decile
Resp_Cnt	No of Observations with Target = 1. Here we are assuming these cases are responders and hence the name Resp_Cnt to indicate count of responders

Rank Order Table Creation

Decile	Base Cnt	#Resp	# Non-Resp	% Resp. Rate	Cum. Base	Cum. Resp.	Cum. Non-Resp	%Cum. Base	%Cum. Resp.	%Cum. Non-Resp	KS
A	B	C	D = B - C	E = C / B	F	G	H	I	J	K	abs (J– K)
10	1,000	295	705	29.5%	1,000	295	705	10%	37.5%	7.7%	29.8%
9	1,000	176	824	17.6%	2,000	471	1,529	20%	59.8%	16.6%	43.3%
8	1,000	115	885	11.5%	3,000	586	2,414	30%	74.5%	26.2%	48.3%
7	1,000	75	925	7.5%	4,000	661	3,339	40%	84.0%	36.2%	47.7%
6	1,000	35	965	3.5%	5,000	696	4,304	50%	88.4%	46.7%	41.7%
5	1,000	30	970	3.0%	6,000	726	5,274	60%	92.2%	57.2%	35.0%
4	1,000	23	977	2.3%	7,000	749	6,251	70%	95.2%	67.8%	27.3%
3	1,000	18	982	1.8%	8,000	767	7,233	80%	97.5%	78.5%	19.0%
2	1,000	13	987	1.3%	9,000	780	8,220	90%	99.1%	89.2%	9.9%
1	1,000	6	993	0.7%	10,000	787	9,213	100%	100.0%	100.0%	0.0%
Total	10,000	787	9,213	7.9%	10,000	787	9,213	100%	100.0%	100.0%	0.0%

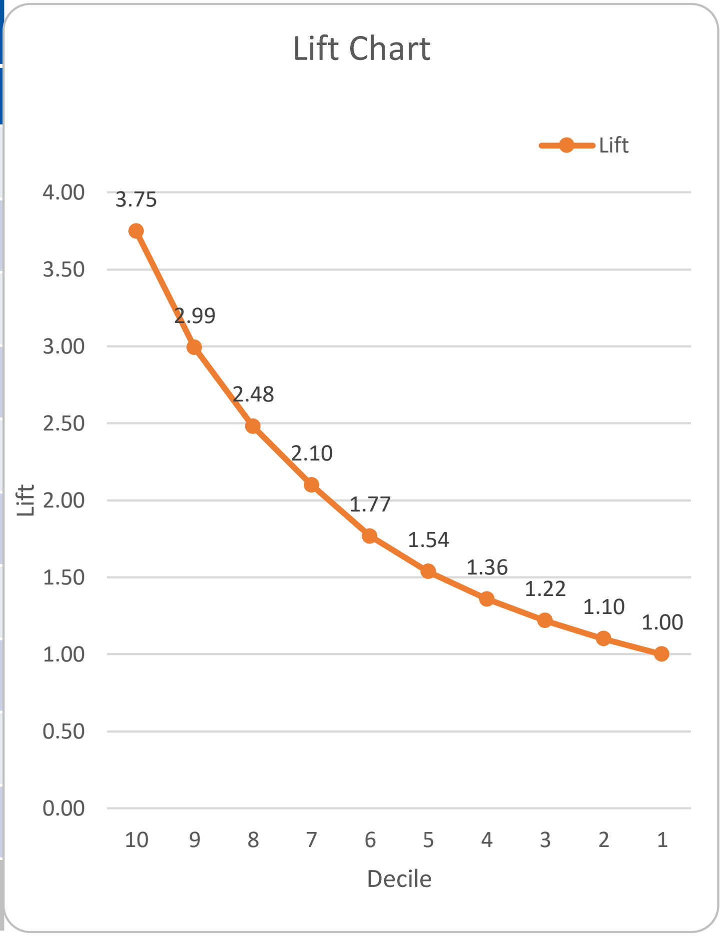
Rank Order Chart

%Cum. Base	%Cum. Resp.	%Cum. Non-Resp
I	J	K
10%	37.5%	7.7%
20%	59.8%	16.6%
30%	74.5%	26.2%
40%	84.0%	36.2%
50%	88.4%	46.7%
60%	92.2%	57.2%
70%	95.2%	67.8%
80%	97.5%	78.5%
90%	99.1%	89.2%
100%	100.0%	100.0%
100%	100.0%	100.0%



Lift Chart

Decile	Base Cnt	#Resp	% Resp. Rate	Cum. Base	Cum. Resp.	Cum. Wt. Resp. Rate	Lift
A	B	C	E = C / B	F	G	G / F	Lift
10	1,000	295	29.50%	1,000	295	29.5%	3.75
9	1,000	176	17.60%	2,000	471	23.6%	2.99
8	1,000	115	11.50%	3,000	586	19.5%	2.48
7	1,000	75	7.50%	4,000	661	16.5%	2.10
6	1,000	35	3.50%	5,000	696	13.9%	1.77
5	1,000	30	3.00%	6,000	726	12.1%	1.54
4	1,000	23	2.30%	7,000	749	10.7%	1.36
3	1,000	18	1.80%	8,000	767	9.6%	1.22
2	1,000	13	1.30%	9,000	780	8.7%	1.10
1	1,000	6	0.70%	10,000	787	7.9%	1.00
Total	10,000	787	7.90%	10,000	787	7.9%	1.00



Concordance

Earning is in Learning
- Rajesh Jakhotia

Concordance – Discordance Calculation

ID	Target	Probability
1	0	0.056
2	0	0.134
3	0	0.156
4	1	0.512
5	0	0.235
6	0	0.250
7	1	0.250
8	1	0.200
9	0	0.135
10	0	0.089

- Concordance Computation Steps:

1. Create Pairs of records such that one record is Target = 1 & other is Target = 0
2. Compare the Pair Probability of Target = 1 w.r.t Probability of Target = 0
3.

3(a). If Prob (T = 1) > Prob (T – 0) then it is **Concordance**

3(b). If Prob (T = 1) < Prob (T – 0) then it is **Dis-Concordance**

3(c). If Prob (T = 1) = Prob (T – 0) then it is **Ties**

4. Concordance = # Concordance Pairs / # Total Pairs

A good model should have Concordance above 75%

Let us create Pairs and compute Concordance

Pairs	Pair Probability	C | D | T
4,1	0.512, 0.056	C
4,2	0.512, 0.134	C
4,3	0.512, 0.156	C
4,5	0.512, 0.235	C
4,6	0.512, 0.250	C
4,9	0.512, 0.135	C
4,10	0.512, 0.089	C

Pairs	Pair Probability	C | D | T
7,1	0.250, 0.056	C
7,2	0.250, 0.134	C
7,3	0.250, 0.156	C
7,5	0.250, 0.235	C
7,6	0.250, 0.250	T
7,9	0.250, 0.135	C
7,10	0.250, 0.089	C

Pairs	Pair Probability	C | D | T
8,1	0.200, 0.056	C
8,2	0.200, 0.134	C
8,3	0.200, 0.156	C
8,5	0.200, 0.235	D
8,6	0.200, 0.250	D
8,9	0.200, 0.135	C
8,10	0.200, 0.089	C

- # **C**oncordance = 18
- # **D**iscordance = 2
- # **T**ies = 1
- # Total Pairs = 21

- % **C**oncordance = 18 / 21 = 85.7%
- % **D**iscordance = 2 / 21 = 9.5%
- % **T**ies = 1 / 21 = 4.8%

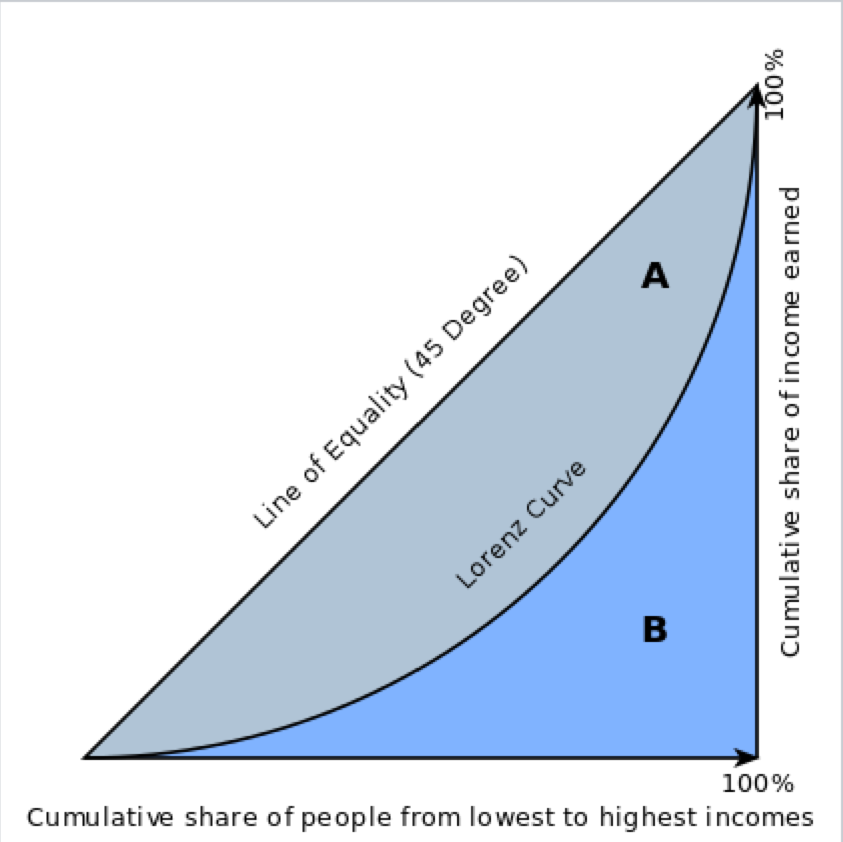
Gini Coefficient

Earning is in Learning
- Rajesh Jakhotia

Lorenz Curve – Inequality of Wealth Distribution

- World's eight richest people have same wealth as poorest 50%

- In economics, the Lorenz curve is a graphical representation of the distribution of income or of wealth. It was developed by Max O. Lorenz in 1905 for representing inequality of the wealth distribution.



A typical Lorenz curve

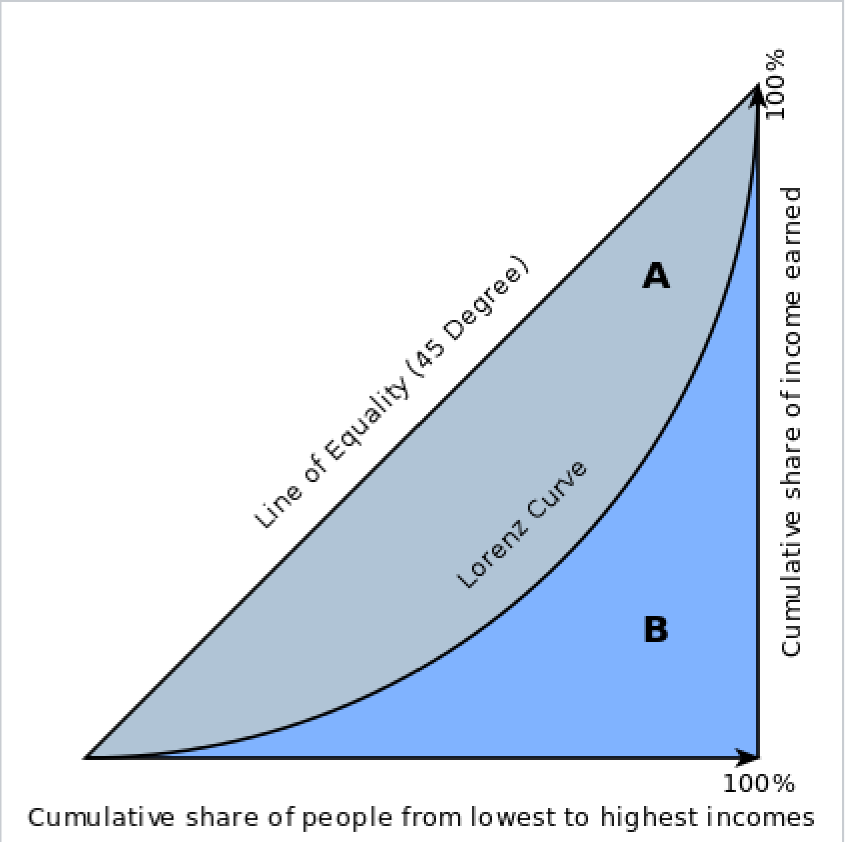
https://en.wikipedia.org/wiki/Lorenz_curve
https://www.theguardian.com/global-development/2017/jan/16/worlds-eight-richest-people-have-same-wealth-as-poorest-50

Lorenz Curve – Inequality of Wealth Distribution

- The Gini Coefficient is the ratio of the area between the line of perfect equality and the observed Lorenz curve to the area between the line of perfect equality and the line of perfect inequality.

- In the diagram on the right, this is given by the ratio *A*/(*A*+*B*), where *A* and *B* are the areas of regions as marked in the diagram.

- The higher the coefficient, the more unequal the distribution is.



A typical Lorenz curve

Confusion Matrix
Classification Error

Earning is in Learning
- Rajesh Jakhotia

Classification Error

- In a Classification Problem, the % of cases wrongly classified is called Classification Error
- Using Predictive Model, we can “Estimate the Probability” Or “Predict the Class”
- Assume we use the Predictive Model to Predict the Class

ID	PV1	PV2	PV3	…	…	…	PVn	(Actual) Target	Predicted Class
1								0	1
2								0	0
.								0	0
.								1	1
.								0	0
								0	0
n								1	1

Confusion Matrix

Confusion Matrix		Predicted Class		Row Total
		0	1	
Actual Target Class	0	True Negative	False Negative (Type 2 Error)	Total Neg.
	1	False Positive (Type 1 Error)	True Positive	Total Pos.

- **Classification Error Rate** is the sum of Type 1 and Type 2 Errors expressed in Percentage terms
- **Sensitive**, also called as True Positive Rate is the proportion of Total Positive that were correctly identified
- **Specificity**, also called as True Negative Rate is the proportion of Total Negatives that were correctly identified

K2 Analytics
Building Skills, Building Individuals

Thank you

Contact us:

ar.jakhotia@k2analytics.co.in