Reading: GLM.Validation
Model: Source Text
Problem Type: Confusion matrix

Given An insurance company wants to make sure its litigation claims get assigned to a senior claims rep as soon as possible.
A logistic model was built to predict the likelihood of a claim going to litigation.

Claim Number	Claim went to Litigation	Predicted Probability of going to Lit
1	Y	96%
2	N	13%
3	Y	37%
4	N	52%
5	N	96%
6	N	21%
7	Y	50%
8	N	28%
9	N	79%
10	Y	91%
11	N	17%
12	Y	91%

Find Calculate confusion matrices for discrimination thresholds of 0.3 and 0.55.

Solution

Claim Number	Claim went to Litigation	Predicted Probability of going to Lit	Discriminant Threshold	
			0.30	0.55
1	Y	0.96	TP	TP
2	N	0.13	TN	TN
3	Y	0.37	TP	FN
4	N	0.52	FP	TN
5	N	0.96	FP	FP
6	N	0.21	TN	TN
7	Y	0.50	TP	FN
8	N	0.28	TN	TN
9	N	0.79	FP	FP
10	Y	0.91	TP	TP
11	N	0.17	TN	TN
12	Y	0.91	TP	TP

Here, TP means True Positive, TN means True Negative, FP means False Positive, and FN means False Negative.

We assign these values as follows:
TP -> Claim went to Litigation = Y and Predicted Probability > Disciminant Threshold
FP -> Claim went to Litigation = N and Predicted Probability > Disciminant Threshold
FN -> Claim went to Litigation = Y and Predicted Probability < Disciminant Threshold
TN -> Claim went to Litigation = N and Predicted Probability < Disciminant Threshold

General Confusion Matrix

		Predicted: Positive	Predicted: Negative
Actual	Positive	TP	FN
	Negative	FP	TN

We assign the count of each type to the matrix.

Threshold = 0.3

5	0
3	4

Threshold = 0.55

3	2
2	5

Observe the lower threshold has less false negatives and more false positives than the higher threshold.
Whether this is a good or bad thing depends on the scenario.
If the price of a false positive is low in terms of money/time/resources and the cost of missing a true positive is high then this is good.
If it is the reverse, i.e. little gain for a lot of cost then this is bad.

Reading: GLM.Validation
Model: Source Text
Problem Type: Confusion matrix

Given An insurance company wants to make sure its litigation claims get assigned to a senior claims rep as soon as possible.
A logistic model was built to predict the likelihood of a claim going to litigation.

Claim Number	Claim went to Litigation	Predicted Probability of going to Lit
1	N	56%
2	Y	38%
3	N	21%
4	N	48%
5	N	48%
6	Y	73%
7	N	14%
8	N	91%
9	Y	80%
10	N	1%
11	N	97%
12	N	97%

Find Calculate confusion matrices for discrimination thresholds of 0.26 and 0.89.

Solution

Claim Number	Claim went to Litigation	Predicted Probability of going to Lit	Discriminant Threshold	
			0.26	0.89
1	N	0.56	FP	TN
2	Y	0.38	TP	FN
3	N	0.21	TN	TN
4	N	0.48	FP	TN
5	N	0.48	FP	TN
6	Y	0.73	TP	FN
7	N	0.14	TN	TN
8	N	0.91	FP	FP
9	Y	0.80	TP	FN
10	N	0.01	TN	TN
11	N	0.97	FP	FP
12	N	0.97	FP	FP

Here, TP means True Positive, TN means True Negative, FP means False Positive, and FN means False Negative.

We assign these values as follows:
TP -> Claim went to Litigation = Y and Predicted Probability > Disciminant Threshold
FP -> Claim went to Litigation = N and Predicted Probability > Disciminant Threshold
FN -> Claim went to Litigation = Y and Predicted Probability < Disciminant Threshold
TN -> Claim went to Litigation = N and Predicted Probability < Disciminant Threshold

General Confusion Matrix

Actual \ Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

We assign the count of each type to the matrix.

Threshold = 0.26

3	0
6	3

Threshold = 0.89

0	3
3	6