|  |  |  |  |
| --- | --- | --- | --- |
| Actual  Predict | 1 (fraudulent) | 0 (non-fraudulent) |  |
| 1(fraudulent) | 30 | 58 | 88 |
| 0(non-fraudulent) | 32 | 920 | 952 |
|  | 62 | 978 | 1040 |

5.1

Error = (32+58) / 1040 = 0.087

5.2

(I treat “The classification error rate” as the percentage that a data has been put in wrong side, otherwise there only has one “Error”: 0.087 could change and I think it only has one trend with the moving down of cutoff: it moves down firstly and moves up lately. In addition, the number of truly fraudulent is always 62, the number of truly non-fraudulent is always 978. Then we need not to describe it separately in a. and b. )

1. The classification error rate for records that are truly fraudulent (Type II),

When the cutoff moving up, the error rate moving up;

When the cutoff moving down, the error rate moving down.

1. The classification error rate for records that are truly non-fraudulent (Type I),

When the cutoff moving up, the error rate moving down;

When the cutoff moving down, the error rate moving up.

(c)

1. "Another analyst comments that you could improve the accuracy of the model by classifying everything as non-fraudulent. If you do that, what is the error rate?"
2. Reconstruct your confusion matrix and recalculate the error rates for this naive classifier's approach to our original Problem 5.1. What happens to the overall error rate? Which error, Type I or II, has this analyst eliminated? Comment also on the profit/cost implications of this analyst's naive classification. Repeat for the case of a perfect classifier.

The chart would be like this after we classify all as non-fraudulent,

|  |  |  |  |
| --- | --- | --- | --- |
| Actual  Predict | 1 (fraudulent) | 0 (non-fraudulent) |  |
| 1(fraudulent) | 0 | 0 | 0 |
| 0(non-fraudulent) | 62 | 978 | 1040 |
|  | 62 | 978 | 1040 |

(I think the error rate here means the Error in complementary classification measures)

The error rate “Error” would be (62+0)/1040 = 0.06.

2)

Overall error rate decreased from 0.087 to 0.06.

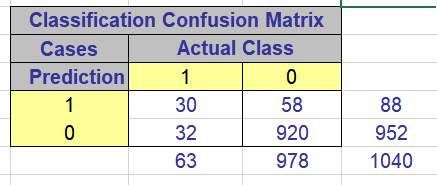
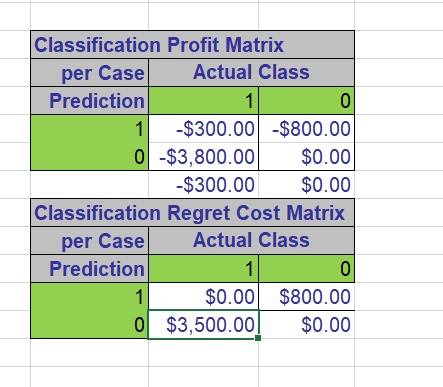
Type I = 0/(0+978) = 0

Type II = 62 / (62 + 0) = 1

This analyst eliminated the Type I error.

This analyst would get all potential profits, but also lose all potential losses. His revenue is rely on his lucky and destiny, too much risk.

5.X



Total profit would be (-300)\*30 + (-800)\*58 + (-3800)\*32 + 0\*920 = -177000 dollars.

Total cost would be 158400 dollars.