Predictive Analytics Lecture 4

Adam Kapelner

Stat 422/722
at The Wharton School of the University of Pennsylvania

February 8 & 9, 2017

Churn Example Where $p_0 = 0.10$

$p_0 = 0.5$		$\hat{y}$ = 1	$\hat{y}$ = 0	Totals	Model Errors
y	1	$TP = 1012$	$FN = 857$	$P = 1869$	$FNR = 45.9\%$
	0	$FP = 531$	$TN = 4632$	$N = 5163$	$FPR = 10.2\%$
	Totals	$\hat{P} = 1543$	$\hat{N} = 5489$	$n = 7032$	
	Use errors	$FDR = 34.3\%$	$FOR = 15.6\%$		$\boxed{ME = 19.7\%}$

$p_0 = 0.1$		$\hat{y}$ = 1	$\hat{y}$ = 0	Totals	Model Errors
y	1	$TP = 1772$	$FN = 97$	$P = 1869$	$FNR = 5.1\%$
	0	$FP = 2669$	$TN = 2494$	$N = 5163$	$FPR = 51.6\%$
	Totals	$\hat{P} = 4441$	$\hat{N} = 2591$	$n = 7032$	
	Use errors	$FDR = 60.1\%$	$FOR = 3.7\%$		$\boxed{ME = 39.3\%}$

Which numbers did not change? n, P and N. Why? These are fixed according to the dataframe. All other numbers changed! What happend to our first means of evaluation, the Misclassification Error? It increased from 19.7% $\rightarrow$ 39.3%. So isn't this a worse model??

Not necessarily... It depends on what your goal is!

Asymmetric Costs in a Classifier

These are always two types of errors but the costs are not always the same.

$p_0 = 0.1$		$\hat{y}$ 1	0	Totals	Model Errors
y	1	$TP = 1772$	$FN = 97$	$P = 1869$	$FNR = 5.1\%$
	0	$FP = 2669$	$TN = 2494$	$N = 5163$	$FPR = 51.6\%$
	Totals	$\hat{P} = 4441$	$\hat{N} = 2591$	$n = 7032$	
	Use errors	$FDR = 60.1\%$	$FOR = 3.7\%$		$\boxed{ME = 39.3\%}$

Imagine we really are the Telecom business manager. It costs 5-10x more to acquire a new customer than to engage a customer who is likely to churn. So you give an incentive package to those who are predicted to churn. Which of the two types of errors specifically is *very* costly? The *FN*. Who are they? These are those who you said were not going to churn *and they did*! Cost? You need to acquire a new customer! The other type of error is less costly, the *FP*. Who are they? These are the people you thought were going to churn and did not. Cost? Whatever the incentive package is.

Weighted Misclassification Error

We now define two costs: (1) the cost of the FP denoted c_{FP} and (2) the cost of the FN denoted c_{FN}. We then define the weighted misclassification error evaluation metric:

$$ME_w := \frac{1}{n} \sum_{i=1}^{n} c_{FP} \mathbb{1}_{y_i=0 \& \hat{y}=1} + c_{FN} \mathbb{1}_{y_i=1 \& \hat{y}_i=0}$$

We now vary p_0 to locate the model that optimizes this error to be minimum.

Minimum Weighted Misclassification Error

Let's assume that $c_{FN} = \$1000$ and $c_{FP} = \$100$ just for the example's sake. Note: this is a **cost ratio** of 10:1.

	Prob	TP	TN	FP	FN	COST
1	0.8117	1	5163	0	1868	1868000
2	0.8104	2	5163	0	1867	1867000
3	0.8093	3	5163	0	1866	1866000
4	0.8092	4	5163	0	1865	1865000
5	0.8090	5	5163	0	1864	1864000
6	0.8085	6	5163	0	1863	1863000
7	0.8083	7	5163	0	1862	1862000
8	0.8082	8	5163	0	1861	1861000
9	0.8079	9	5163	0	1860	1860000

We now calculate the cost and find the minimum model (i.e. the p_0 to ship). [JMP] Or alternatively, we can select the model with the closest $FN/FP \approx 10:1$ to match the stakeholder preference of the desired cost ratio. Why would this be good?

Expected Value Calculation

You can also imagine assignment of both costs *and* benefits:

$p_0 = 0.1$		$\hat{y}$ 1	0
y	1	b_{TP}	c_{FN}
	0	c_{FP}	b_{TN}

and then use the confusion matrix to estimate probabilities:

$p_0 = 0.1$		$\hat{y}$ 1	0
y	1	25.1%	1.3%
	0	40.0%	35.5%

The expected value would be?

$$\begin{aligned} \mathbb{E}[T] &= p_{TP} \times b_{TP} + p_{TN} \times b_{TN} + p_{FP} \times c_{FP} + p_{FN} \times c_{FN} \\ &\approx \hat{p}_{TP} \times b_{TP} + \hat{p}_{TN} \times b_{TN} + \hat{p}_{FP} \times c_{FP} + \hat{p}_{FN} \times c_{FN} \end{aligned}$$

Highest expected value model is shipped (ex. from Provost & Fawcett, 2013).

$\hat{p}$'s as Ordinal Values

One final point... If we were on a mission to find the top m churners. What would we do? Sort the $\hat{p}$'s and return the top m.