Lecture 21 04/23/2020

$y = \{0, 1\}$ model binary classification or prob. est

Two types of error

① $\hat{y} = 0, y = 1$ False Negative (FN). cost FN (C_{FN})

② $\hat{y} = 1, y = 0$ False Positive (FP). cost FP (C_{FP})

There are costs to these errors:- C_{FP} & C_{FN}

If $C_{FP} \neq C_{FN}$ this is called asymmetric costs.

Asymmetric cost Classification: a classification model that attempts to minimize total cost. while initializing both C_{FP}, C_{FN} which are different

y \ $\hat{y}$	0	1	
0	TN	FP	N (# of neg.)
1	FN	TP	P (# of pos.)
# of predict Neg. & posi.	PN	PP	n

Most important

False Discover Rate $FDR := \frac{FP}{PP} = 1 - \text{precision}$

False Omission Rate $FOR := \frac{FN}{PN}$

False Positive Rate $FPR := \frac{FP}{\cancel{P}N}$

misclassification error

(error rate)

$$\text{error} = \frac{FP + FN}{n}$$

(accuracy)

$$\text{acc} := 1 - \text{error} = \frac{TP + TN}{n}$$

$$\text{precision} := \frac{TP}{PP}$$

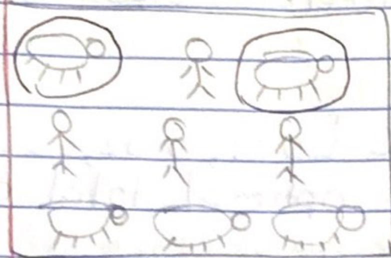
$$\text{recall (sensitivity)} := \frac{TP}{P}$$

$$F_1 := \frac{2}{1/\text{precision} + 1/\text{Recall}}$$

which may not be defined.

Identity bags  = 1 precision = $\frac{2}{2}$ = 100%

 = 0



recall = $\frac{2}{5}$ = 40%

		$\hat{y}$ 0	1	
y	0	4	0	4
	1	3	2	5
		7	2	9

error = $\frac{3}{9}$ = 33%

$F_1 = \frac{2}{\frac{1}{1} + \frac{1}{0.4}} = 0.57 = 57\%$

Overall cost: $C = C_{FP} FP + C_{FN} FN$ (we want to minimize this)

Overall Reward: $R = C + \underbrace{r_{TP} TP + r_{TN} TN}_{}$

Stakeholders-specified reward (>0)

What is an algorithm that will allow for asymmetric cost models?

Remember logistic regression. This is a probability estimation model.

$\Rightarrow \hat{p} = g_{pr}(\vec{x}_*) = \hat{p}(Y_* = 1 | \vec{x}_*) \notin [0,1]$. This is not $\hat{y}$. It is not classification.

How can we "rig" prob. est. models to become classification models?

$$\hat{y} = \mathbb{1}_{\hat{p} \geq 0.5} = \begin{cases} 1 & \text{if } \hat{p} \geq 0.5 \\ 0 & \text{o/t} \end{cases}$$

hyperparameter in $\mathcal{A}$

What if $\hat{y} = \mathbb{1}_{\hat{p} \geq 0.9}$ → reduce PP → reduce FP
→ increase PN → increase FN

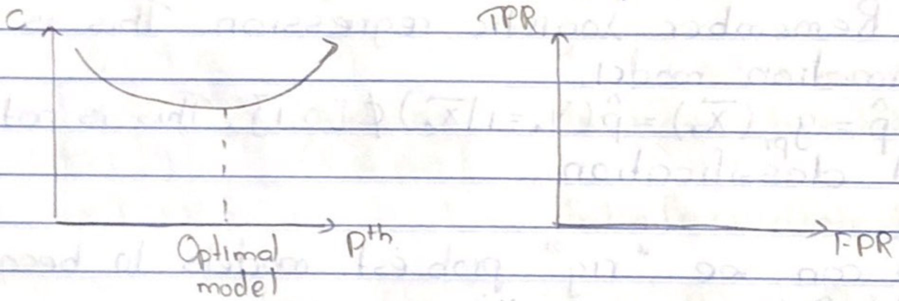
What if $\hat{y} = \mathbb{1}_{\hat{p} \geq 0.1}$ → reduce PN → reduce FN
→ increase PP → increase FP

In general $\hat{y} = \mathbb{1}_{\hat{p} \geq p_{th}}$ — threshold hyperparameter. Each unique value in a different classification model.

We choose a grid path $\in \{0.01, 0.02, \dots, 0.99\}$ then p^{th} based ~~based~~ on optimal total cost | total reward.

p^{th}	TP	TN	FP	FN	precision	recall	FDR	FOR	error	F_1	C	R
0.01												
0.02												
⋮												
0.98												
0.99												

There are a number of popular illustritions.



Recall, sensitivity TPR (True Positive Rate) $= \frac{TP}{P}$

$$FPR = \frac{FP}{N} = 1 - \text{specificity}.$$

Purely Random Model w/ p^{th} := $\hat{p}$ is realization from $U(0,1)$ call it u.

$$\hat{y} = \mathbb{1}_{\hat{p} \geq p^{th}} = \mathbb{1}_{u \geq p^{th}}$$

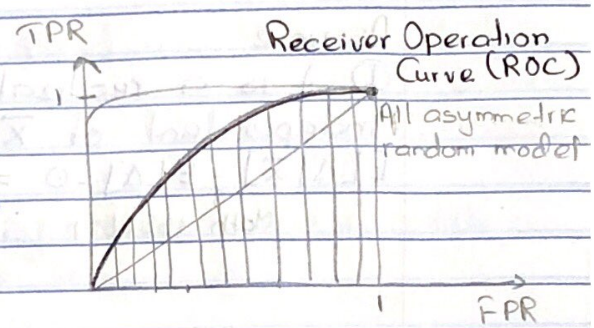
		$\hat{y}$		
		0	1	
y	0	$nP_{th}(1-\bar{y})$	$n(1-P_{th})\cdot(1-\bar{y})$	$n(1-\bar{y})$
	1	$nP_{th}\bar{y}$	$n(1-P_{th})\bar{y}$	$n(\bar{y})$
		nP_{th}	$n(1-P_{th})$	n

$P_y = \bar{y}$

→ Fixed quantity regardless of pth

$$TPR = \frac{n(1-P_{th})\bar{y}}{n\bar{y}} = 1 - P_{th}$$

$$FPR = \frac{n(1-p_{th})(1-\bar{y})}{n(1-\bar{y})} = 1 - P_{th}$$



Q want an overall metric about how my prob. est. model is doing

AUC: Area under RAC curve $\in [0,1]$ really $[0.5, 1]$

Detection Error Tradeoff (DET) Plot