$y = \{0,1\}$ Model: binary classification or prob_est.

Two types of errors.

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 internalizing both c_{FP}, c_{FN} which are different.

		$\hat{y}$: 0	1		
y	0	TN	FP	N	(# negative)
	1	FN	TP	P	(# positive)
# predicted neg & pos:		PN	PP	n	

(error rate) misclassification error
$$err := \frac{FP+FN}{n}$$

accuracy
$$acc := 1 - err = \frac{TP+TN}{n}$$

false discovery rate $FDR := \frac{FP}{PP} = 1 - \text{precision}$

false omission rate $FOR := \frac{FN}{PN}$

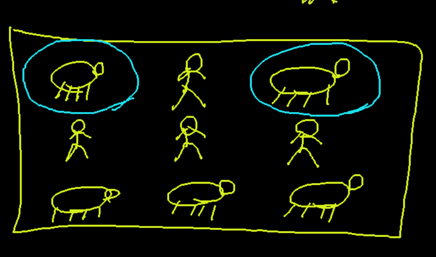
$$\text{precision} := \frac{TP}{PP}$$

Sensitivity / recall $:= \frac{TP}{P}$

$$F_1 := \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$ which may not be defined

False positive rate $FPR := \frac{FP}{N}$

Identify bugs
= 1
= 0



$$\text{precision} = \frac{2}{2} = 100\%$$

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

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

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

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

* Overall Cost: $C = c_{FP} FP + c_{FN} FN$. We wish to minimize this.

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

stakeholder-specified rewards (≥ 0).

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

Remember logistic regression. This is a probability estimation model.

$\Rightarrow \hat{p} = g_{pr}(\vec{x}_*) = \hat{P}(Y_* = 1 \mid \vec{x}_*) \notin \{0,1\}$. 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 $\hat{y} = \mathbb{1}_{\hat{p} \geq 0.9}$ $\Rightarrow$ reduces PP $\Rightarrow$ reduces FP
$\Rightarrow$ increases PN $\Rightarrow$ increases FN

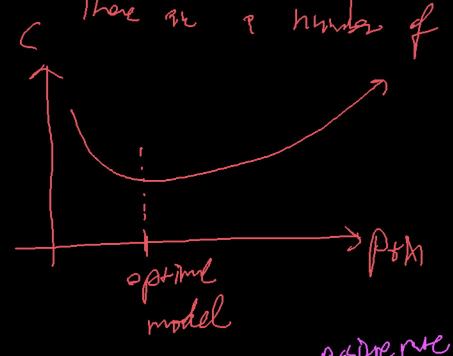
'' '' $\hat{y} = \mathbb{1}_{\hat{p} \geq 0.1}$ $\Rightarrow$ reduces PN $\Rightarrow$ reduces FN
$\Rightarrow$ increases PP $\Rightarrow$ increases FP

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

We choose a grid $p_{th} \in \{0.01, 0.02, \dots, 0.99\}$ then p_{th} based on optimal total cost / total reward.

p_{th}	TP	TN	FP	FN	precision	recall	FDR	FOR	err	F_1	C	R
0.01												
0.02												
⋮												
0.98												
0.99												

There are a number of popular illustrations



true positive rate

recall, sensitivity, $TPR = \frac{TP}{P}$

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

Purely Random Model with p_{th}: $\hat{p}$ is a realization from $U(0,1)$ call it u.

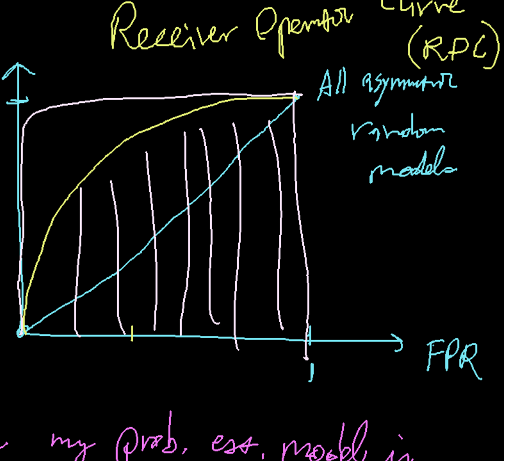
$$\hat{y} = \mathbb{1}_{\hat{p} \geq p_{th}} = \mathbb{1}_{u \geq p_{th}}$$

$p_y = \bar{y}$

		$\hat{y}$: 0	1	
y	0	$n p_{th}(1-\bar{y})$	$n(1-p_{th})(1-\bar{y})$	$n(1-\bar{y})$
	1	$n p_{th} \bar{y}$	$n(1-p_{th})\bar{y}$	$n\bar{y}$
		$n p_{th}$	$n(1-p_{th})$	n

} fixed quantities regardless of p_{th}.

Receiver Operator Curve (ROC)



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

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

I want an overall metric about how my prob. est. model is doing. AUC: area under ROC curve $\in [0,1]$ really $[0.5, 1]$

Detection Error Tradeoff Plot (DET).