Asymmetric cost modeling for binary classification

$y = \{0, 1\}$

There are two possible errors in binary classification models.

I yhat = 0 but y = 1 (false negative, FN)

II yhat = 1 but y = 0 (false positive, FP)

Consider the scenario where the "cost" of FN (c_FN) is not the same as the "cost" of FP (c_FP). Adam's example:
modeling a fire in a building i.e. with an alarm (y=1 is there's a fire).
FP: if you tell the resident they have a fire
and the reality is they don't have a fire
FN: if you tell the resident they don't have a fire
and the reality is they do have a fire.
"Assymetric costs" means that c_FN > c_FP or c_FN < c_FP.
In this case c_FN > c_FP!! c_FN is major destruction and possible death and c_FP is you take your towel and go like this.

Here is the 2x2 confusion matrix:

$y \backslash \hat{y}$	0	1	
0	TN	FP	N
1	FN	TP	P
	PN	PP	n

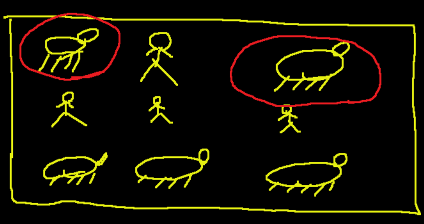
TN: true negatives, TP: true positives,
FN: false negatives, FP: false positives,
N: # of negatives, P: # of positives,
PN: # of predicted negatives,
PP: # of predicticed positives,
n: sample size

$$ME := \frac{FN + FP}{n}, \quad ACC := 1 - ME = \frac{TN + TP}{n}$$

precision := $\frac{TP}{PP}$ what proportion of your positive predictions are correct?

recall := $\frac{TP}{P}$ what proportion of the positives did you locate?

$y = 1 \Rightarrow$ BUG



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

F_1 := $\frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = .57$ a way of averaging precision and recall

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

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

There's a tradeoff between precision and recall AND there's a tradeoff between false discovery rate and false omission rate (FDR vs FOR)

Another metric uses the assymmetric costs c_FN and c_FP. Before doing this, you need to assign values to the costs. For example, in the fire alarm, maybe c_FP = 1 and c_FN = 500. Imagine it in dollars. A natural error metric now is "total cost"

$$* \; C := c_{FP} FP + c_{FN} FN$$

$$W := w_{TN} TN + w_{TP} TP + c_{FP} FP + c_{FN} FN$$

What if I want to tailor my model to minimize the cost? This means that the model must incorporate c_FN and c_FP.

What if I have a probability estimation model and I want to use it for classification. How do I do that?

$$\hat{p} \longrightarrow \hat{y} \quad \text{e.g.} \quad \hat{y} = \mathbb{1}_{\hat{p} \geq 50\%} \underset{\text{general}}{=} \mathbb{1}_{\hat{p} \geq p_{th}}$$

threshold hyperparam (default is 50%)

What if p_th = 90%? $\hat{y} = \mathbb{1}_{\hat{p} \geq 90\%}$

$c_{FP} > c_{FN} \Rightarrow p_{th}$ high

$y \backslash \hat{y}$	0	1
0		FP LOW
1	FN HIGH	
	PN	PP
	HIGH	LOW

p_th is a tuning hyperparameter
which is tuned based on the
assymmetric cost values. But...
how do we pick a p_th to use in the predictive model g?

Each unique value of p_th will result in a different model g.
You need to do model selection. We will pick the p_th
that results in lowest total cost. First step is to make a grid:

$$p_{th} \in [0, 1] \qquad p_{th} \in \{0.01, 0.02, \ldots, 0.99\}$$

Then you let the computer fill in the following table:

p_{th}^*	TP	TN	FP^*	FN^*	precision	recall	FDR	FOR	$C^* = c_{FP}FP + c_{FN}FN$
0.01									
0.02									
⋮									
0.99									

Optimal model