13) More Performance Metrics

Confusion Matrix & APRF

Model \ Reality	Present	Absent
Yes	True positive	False positive
No	False negative	True negative

shows all the ways for M to get confused with R

The Metrics

① Accuracy $= \frac{TP+TN}{\#\text{ samples}}$ (overall model performance; ignores category size)

② Precision $= \frac{TP}{\#\text{ yes}}$ (Penalizes FP – wrong yes; Useful for disease detection)

③ Recall $= \frac{TP}{\#\text{ present}}$ (Penalizes FN – wrong no; Useful for covid-19 test)

④ F1 $= \frac{TP}{TP+\text{avg}(FP+FN)}$ (high means model is unbiased; Balance b/t ② & ③)

Example: Given CM $\begin{bmatrix} 70 & 5 \\ 20 & 5 \end{bmatrix}$, calculate the 4 metrics above

$A = \frac{70+5}{100} = 0.75$, $P = \frac{70}{70+5} = 0.933$, $R = \frac{70}{70+20} = 0.778$, $F1 = \frac{70}{70+\frac{5+20}{2}} = 0.848$

Model has biases for saying no & is wrong overall

Performance: $S_{train} < S_{test}$?

- **Reason:** i) Test data are easier to categorize (systematic bias, lack randomness)
  ii) Bug in code (ex. test data taken from training)
- **Solution:** If big differences → i) Debug ii) simplify model architecture