Classification Metrics

→ Logistic Regression
→ Confusion Matrix
  ↳ Precission
  ↳ Recall
→ F1-score

Accuracy ⇒ $\frac{\#\ \text{correctness}}{\#\ \text{total}}$

Model → 90% Accuracy

Spam Detector (Gmail)

Inbox (Ham) ← 90%
Spam ← 10%

Model →

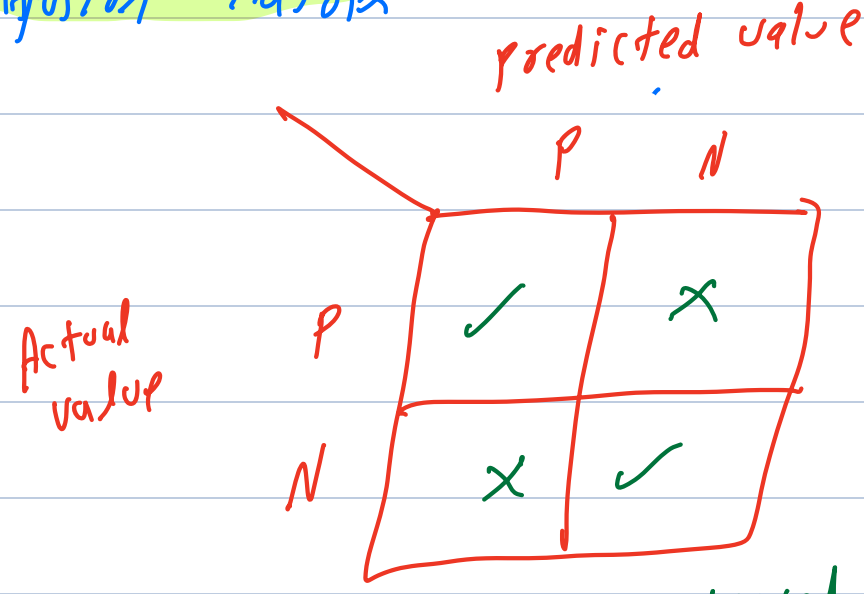
New Email → Ham
100 email → Ham ✓

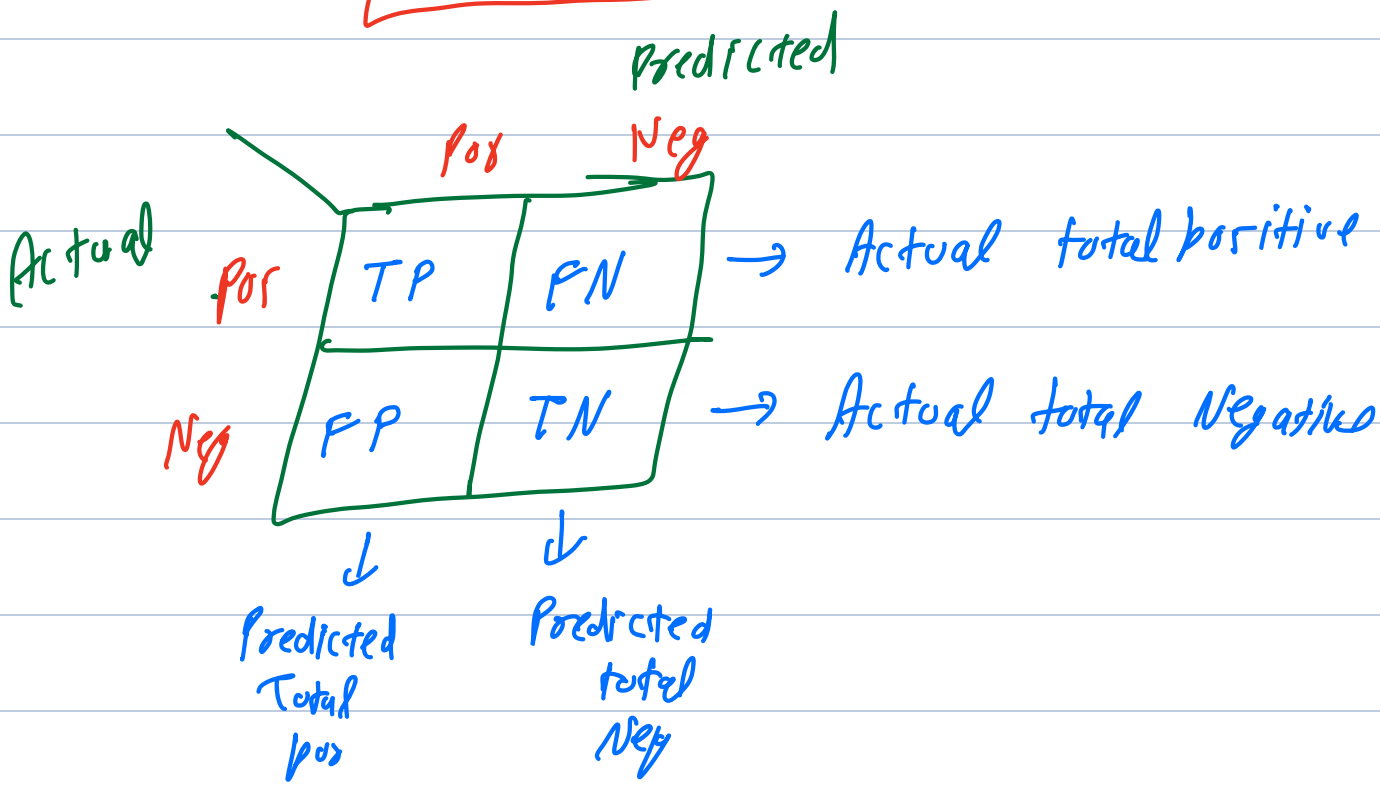
90% → correct 10%

↓↓

Dumb model

Confusion Matrix





⇒ Predicted mail as spam & it actual is spam

TP

⇒ Predicted mail as not spam & it's actually a spam

TN

⇒ Predicted mail as not spam & it actually is a spam

FN → Type 2 error

⇒ Predicted mail as not spam & it is a spam

(FN)

⇒ 500 → 100 → 5
→ 400 → 400

$$\frac{5+400}{500} \approx 0.81$$

⇒ Precision & Recall

Fishing → water

100 fishes → positive

100 turtles → Negative

Exactnes

60 fishes & 20 turtles

$\frac{60}{60+20}$

Recall $\frac{60}{100}$

E+D →

80 F & 80 T

Precision ⇒ $\frac{80}{160}$

Recall → $\frac{80}{100}$

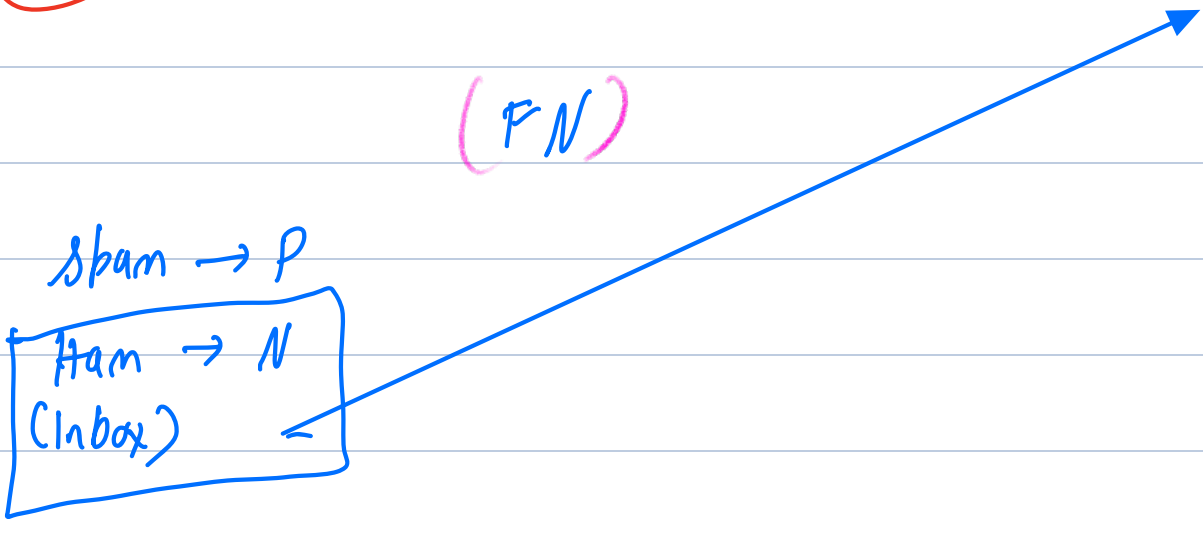
Ex3 ⇒ 20F & No turtle

Precision → $\frac{20}{20} = 100\%$

Recall ⇒ $\frac{20}{100} \Rightarrow 20\%$

Precision

① Recieve an spam email in inbox



② appreciation Email from manager → spam

FP

TP ↑ ⇒ Metric ↑ FP ↑ ⇒ Metric ↓

$\frac{TP}{\text{Predicted}} \Rightarrow \frac{TP}{TP + FP} \simeq$ Precision

total Positive

$$\frac{\text{Fish}}{\text{Fish + Turtle}} \Rightarrow \frac{\text{TP}}{\text{Predicted total Pos}}$$

Recall ⇒

Positive → cancer

Negative → Non cancer

① Classify a healthy person as having cancer

FP

② Misclassify a person having cancer as healthy

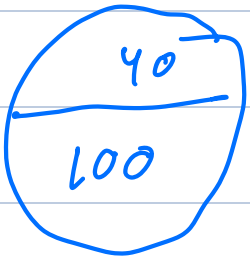
FN

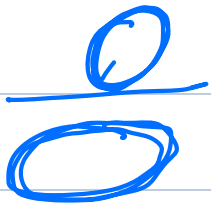
TP ↑ FN ↓

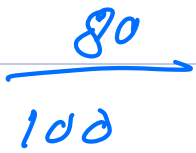
$$\text{Recall} \simeq \frac{TP}{TP + FN} = \frac{TP}{\text{Total Actual Pos}}$$

$$\simeq \frac{60 \text{ Fish}}{100 \text{ Fish}}$$

⇒ Movie Recommendation







Its okay to miss few good recommendation

⇒ detect Fraud txn

Non Fraud → Fraud (FP) ✗

Fraud → Non Fraud (FN) ✓
↘ Recall