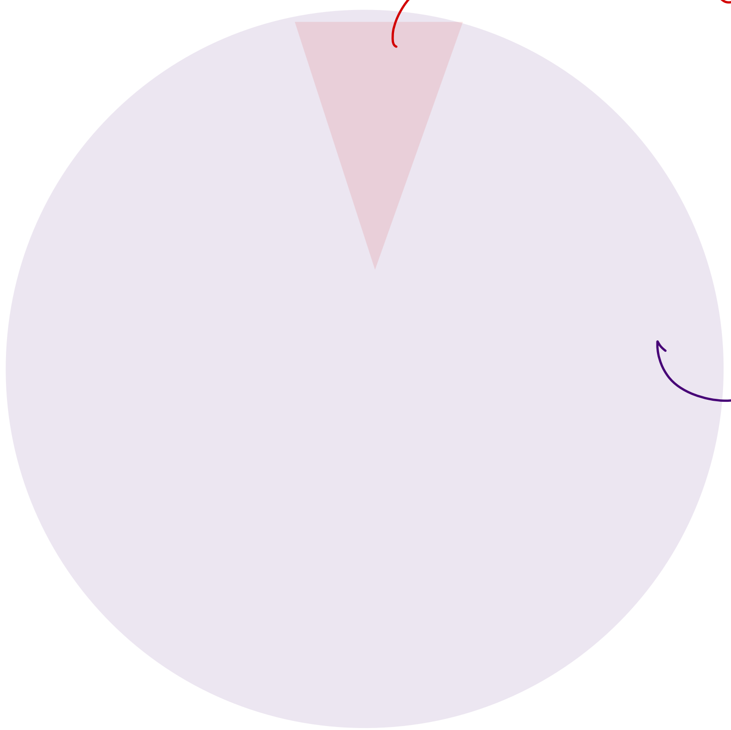
Model Evaluation and Selection

- Evaluation metrics
  - How can we measure accuracy?
  - Other metrics to consider?
- Use **validation test set** of class-labeled tuples instead of training set when assessing accuracy
- Methods for estimating a classifier's accuracy
  - Holdout method
  - Cross-validation
  - Bootstrap
- Comparing classifiers:
  - ROC Curves

→ Positive

2% classifier → ไม่ได้ผล (ความแม่นยำ 98%)



อยู่ 98% → Neg

ถ้า Test data

Model ทาย neg อย่างเดียว → 100 → Pos 2

Neg 98

→ Accuracy = 98%

Precision $\frac{0}{0+0}$: undefined , Recall $\frac{0}{0+2} = 0$

specificity $\frac{TN}{N} = \frac{98}{98} = 100\%$

Transpost

ความจริง \ ทำนาย	P	N
P	TP: 0	FP: 0
N	FN: 2	TN: 98

ทำนาย \ ความจริง	P	N
P	TP: 0	FN: 2
N	FP: 0	TN: 98

จากผลที่ได้ใน python ; array([[30, 6]
[2, 48]])

0 = malig (เนื้อร้าย) pos , 1 = Benign neg

Sens - 0.83

Spec - $\left(\frac{48}{48+2}\right)$

1 เป็น pos

TN 0 | 1 FP

0 [30, 6]

1 [2, 48]

FN | TP

0 เป็น pos

จริง \ ทำนาย	TP 0	1 FN
0	30	6
1	2	48
	FP	TN

ทำนายว่าเป็นจริง ถูกมากแค่ไหน

class ที่เป็น pos จริง

สนใจผล positive (ให้ 1 เป็น positive)

ให้ 0 เป็น positive

ให้ 1 เป็น Positive

ทำนายว่าเป็น pos ถูกมากแค่ไหน

ความที่เป็น pos จริง

ค่าเฉลี่ยของ precision, recall

	precision	recall	f1-score	support
0	0.94 ($\frac{30}{30+2}$)	0.83 ($\frac{30}{30+6}$)	0.88	36
1	0.89 ($\frac{48}{48+6}$)	0.96 ($\frac{48}{48+2}$)	0.92	50
accuracy			0.91 ($\frac{48+30}{86}$)	86
macro avg (ค่าเฉลี่ยปกติ)	0.91	0.90	0.90	86
weighted avg	0.91	0.91	0.91	86

เฉลี่ย P และ R

ในข้อสอบที่ถามถูกเท่าไหร่ $\left(\frac{TN+TP}{All}\right)$

ค่าถ่วงน้ำหนักของค่าเฉลี่ย

$$\frac{(0.83\times36)+(0.96\times50)}{36+50}$$

Classifier Evaluation Metrics: Confusion Matrix

- **Confusion Matrix:**

Actual class\Predicted class	C_1	$\neg C_1$
C_1	**True Positives (TP)**	**False Negatives (FN)**
$\neg C_1$	**False Positives (FP)**	**True Negatives (TN)**

- In a confusion matrix w. *m* classes, $CM_{i,j}$ indicates # of tuples in class *i* that were labeled by the classifier as class *j*
  - May have extra rows/columns to provide totals
- **Example of Confusion Matrix:**

Actual class\Predicted class	buy_computer = yes	buy_computer = no	Total
buy_computer = yes	**6954**	**46**	7000
buy_computer = no	**412**	**2588**	3000
Total	7366	2634	10000

Classifier Evaluation Metrics: Accuracy, Error Rate, Sensitivity and Specificity

A\P	C	¬C	
C	**TP**	**FN**	**P**
¬C	**FP**	**TN**	**N**
	P'	**N'**	**All**

- **Classifier accuracy,** or recognition rate
  - Percentage of test set tuples that are correctly classified

  Accuracy = (TP + TN)/All
- **Error rate:** *1 – accuracy*, or

  Error rate = (FP + FN)/All

- **Class imbalance problem**
  - One class may be *rare*
    - E.g., fraud, or HIV-positive
  - Significant *majority of the negative class* and minority of the positive class
  - Measures handle the class imbalance problem
    - **Sensitivity** (recall): True positive recognition rate
      - **Sensitivity = TP/P**
    - **Specificity**: True negative recognition rate
      - **Specificity = TN/N**