Model Evaluation

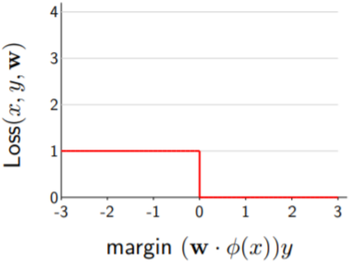
Ninghao Liu

University of Georgia

February 1, 2023

Some contents adopted from “Data Mining”, Section 8.5.

Where we left off: Classification Error

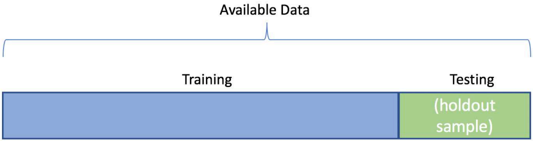


$$L_{0-1}(x, y, \boldsymbol{w}) = \mathbb{1}[y' \neq y] = \mathbb{1}[(\boldsymbol{w}^\intercal \cdot \phi(x))y \leq 0]$$

We define **Accuracy** as:

$$Acc = \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{x},y)\in\mathcal{D}} \mathbb{1}[y' = y] = \frac{\#\text{correct predictions}}{\#\textit{predictions}} \tag{1}$$

Training, Validation and Testing in Practice



To report model performance:

- $\mathcal{D}_{in}$ (Training set): totally contaminated.
- $\mathcal{D}_{test}$ (Test set): totally clean.

Questions

1. **How to choose $\mathcal{D}$?**
    - $\mathcal{D}_{test}$ in testing to report final performance.
    - $\mathcal{D}_{val}$ in model selection and hyper-parameter tuning.
2. Are these evaluation metrics enough?

Questions

1. How to choose $\mathcal{D}$?
2. **Are these evaluation metrics enough?**

Questions

Scenarios:

- Bomb detection.
    - $+$: Bomb detected. $-$: No bomb.
- Email spam detection.
    - $+$: Spam email. $-$: Normal email.

Is accuracy a good metric?

Evaluation Terminology

In the classification scenario (especially multi-class classification):

- **Positive samples** (positive tuples)
  - Tuples of the main class of interest
- **Negative samples** (negative tuples)
  - All other tuples

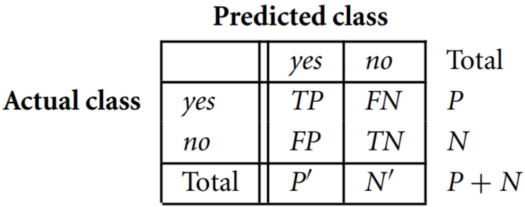
P is the number of positive tuples and N is the number of negative tuples.

Evaluation Terminology

For each tuple, compare the predictor's class label prediction with the tuple's ground-truth class label.

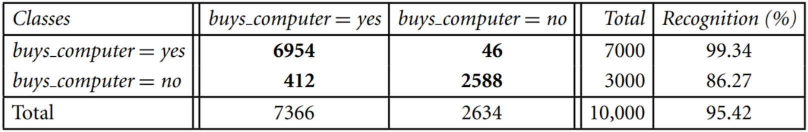
- **True positives** (*TP*)
  - The *positive* tuples that were *correctly* predicted.
  - Let *TP* be the number of true positives.
- **True negatives** (*TN*)
  - The *negative* tuples that were *correctly* predicted.
  - Let *TN* be the number of true negatives.
- **False positives** (*FP*)
  - The negative tuples that were *incorrectly* predicted (as positive).
  - Let *FP* be the number of false positives.
- **False negatives** (*FN*)
  - The positive tuples that were *incorrectly* predicted (as negative).
  - Let *FN* be the number of false negatives.

Confusion Matrix



		Predicted class		
		yes	*no*	Total
Actual class	*yes*	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	$P + N$

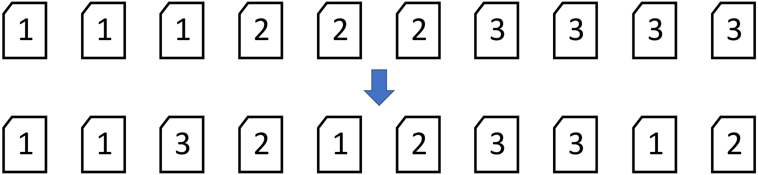
Confusion Matrix



Classes	*buys_computer = yes*	*buys_computer = no*	*Total*	*Recognition (%)*
buys_computer = yes	**6954**	**46**	7000	99.34
buys_computer = no	**412**	**2588**	3000	86.27
Total	7366	2634	10,000	95.42

Confusion Matrix

Given C classes (where $C \geq 2$), a confusion matrix is a table of at least size C by C.



Commonly Used Metrics

- accuracy: $\frac{TP+TN}{P+N}$
- error rate: $\frac{FP+FN}{P+N}$

- True positive rate (TPR): $\frac{TP}{P}$
  - Also called "sensitivity".
- True negative rate (TNR): $\frac{TN}{N}$
  - Also called "specificity".

$$accuracy = sensitivity\frac{P}{P+N} + specificity\frac{N}{P+N}$$

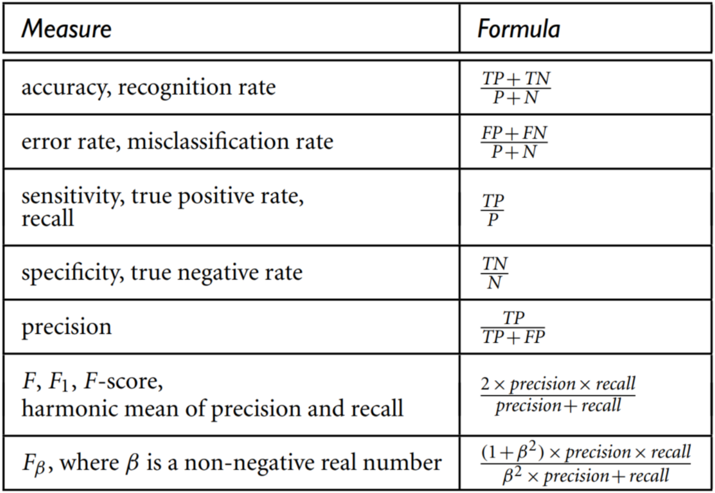
Commonly Used Metrics

- Precision: $\frac{TP}{TP+FP}$
  - what percentage of tuples predicted as positive are actually such
- Recall: $\frac{TP}{TP+FN} = \frac{TP}{P}$
  - what percentage of positive tuples are predicted as such

Classification Error

- F measure: $\frac{2 \times \mathit{precision} \times \mathit{recall}}{\mathit{precision} + \mathit{recall}}$
  - The harmonic mean of precision and recall
  - Also known as F_1 **score** or F-score
- F_β measure: $\frac{(1+\beta^2) \times \mathit{precision} \times \mathit{recall}}{\beta^2 \times \mathit{precision} + \mathit{recall}}$

Confusion Matrix



Measure	*Formula*
accuracy, recognition rate	$\frac{TP+TN}{P+N}$
error rate, misclassification rate	$\frac{FP+FN}{P+N}$
sensitivity, true positive rate, recall	$\frac{TP}{P}$
specificity, true negative rate	$\frac{TN}{N}$
precision	$\frac{TP}{TP+FP}$
F, F_1, F-score, harmonic mean of precision and recall	$\frac{2 \times precision \times recall}{precision + recall}$
F_β, where β is a non-negative real number	$\frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$