

Universitat Politècnica de Catalunya · Barcelona
Tech - UPC
Escola d'Enginyeria de Barcelona Est EEBE
Homework 2, Metrics for Evaluating Multi-Class
Classification Problems

Sheik, Abdullahi

October 2023

Problem 1: Paper [1] defines the average accuracy (equation (1) in the paper) as the sum of the true positive rate of each class divided by the number of classes. Is the same expression as in slide 19?

1. The first equation calculates the average accuracy by taking the average of the True Positive Rate (or Recall) of each class. The True Positive Rate for a class is calculated as:

$$\text{TPR}_i = \frac{\text{tp}_i}{\text{tp}_i + \text{fn}_i} \quad (1)$$

where TPR_i is the True Positive Rate for class i , tp_i is the number of True Positives for class i , and fn_i is the number of False Negatives for class i .

$$\text{acc} = \frac{1}{C} \sum_{i=1}^C \text{TPR}_i \quad (2)$$

where C is the number of classes.

2. The second equation calculates the average accuracy as the average of the sum of True Positives and True Negatives divided by the total number of instances (True Positives, True Negatives, False Positives, False Negatives) for each class:

$$\overline{\text{acc}} = \frac{1}{l} \sum_{i=1}^l \frac{\text{tp}_i + \text{tn}_i}{\text{tp}_i + \text{fn}_i + \text{fp}_i + \text{tn}_i} \quad (3)$$

where l is the number of classes.

Summary

- The first equation calculates the average of the True Positive Rates (Recall) across all classes, considering only True Positives and False Negatives for each class.
- The second equation calculates the average accuracy by considering all True Positives, True Negatives, False Positives, and False Negatives for each class.

They measure different aspects of a classifier's performance and are not equivalent.

1 Problem 2: Paper [2] defines the accuracy and the macro-average arithmetic, as well as the partial accuracy of each class. How do these three measures relate to the average accuracy in paper [1]?

Let's break down the provided equations to understand them better.

Equation 1:

$$\text{Acc} = \frac{\sum_{i=1}^m \sum_{j=1}^c f(i, j) \cdot C(i, j)}{m} \quad (4)$$

- Acc: Represents the accuracy.
- m : The total number of instances.
- c : The number of classes.
- $f(i, j)$: It might represent a binary indicator reflecting the correct classification of an instance, i.e., it is 1 if instance i is correctly classified as class j and 0 otherwise.
- $C(i, j)$: Represents a cost function or a confusion matrix element, depending on the context.

This equation sums over all instances and all classes and divides by the total number of instances. It seems to represent an overall accuracy, weighted by the cost function or confusion matrix element.

Equation 2:

$$\text{MAvA} = \sum_{j=1}^c \frac{\left(\frac{\sum_{i=1}^m f(i,j) \cdot C(i,j)}{m_j} \right)}{C} \quad (5)$$

- MAvA: Represents the macro-average accuracy.
- m_j : Represents the number of instances in class j .
- C : The number of classes.

This equation calculates a per-class average (hence the term "macro-average"), where for each class, it calculates a weighted average (by $f(i,j) \cdot C(i,j)$) and then averages over all classes.

The first equation, Acc, is an overall measure, likely a form of weighted accuracy where each instance's contribution is weighted by the corresponding $C(i,j)$.

The second equation, MAvA, is a form of macro-averaged accuracy where each class contributes equally to the final average, regardless of the number of instances in each class.

In summary, while both equations incorporate weights and involve summation over instances and classes, they represent different forms of averages: the first one is more of a global, weighted average, and the second one is a macro-average where each class has an equal contribution to the final average, irrespective of its size (number of instances).

Problem 3: Paper [3] also defines the average accuracy. How does this measure relate to the measures in papers [1] and [2]? Compute all these measures for the table in Problem 1.

Given the corrected confusion matrix:

	A	B	C	D
A	2542	2	1	15
B	7	1231	40	2
C	1	45	1230	4
D	36	6	3	1235

Table 1: Confusion Matrix

Corrected Metrics for Each Class:

- **True Positives (tp):** A: 2542, B: 1231, C: 1230, D: 1235

- **False Positives (fp):** A: 44, B: 53, C: 44, D: 21
- **False Negatives (fn):** A: 18, B: 49, C: 50, D: 45
- **True Negatives (tn):** A: 3786, B: 5057, C: 5066, D: 5089
- **True Positive Rate (Recall) (TPR):** A: 0.993, B: 0.962, C: 0.961, D: 0.965
- **Precision (ppv):** A: 0.983, B: 0.959, C: 0.965, D: 0.983
- **Specificity (tnr):** A: 0.989, B: 0.990, C: 0.991, D: 0.996
- **Accuracy (per class):** A: 0.990, B: 0.984, C: 0.985, D: 0.990

Corrected Overall Metrics:

- **Average Accuracy:** 0.987
- **Macro Average Accuracy (MAvA):** 0.987
- **Weighted Accuracy (Acc) from problem 2:** 1724.571