Micro- and macro-accuracy

In the final homework assignment, you were asked to compute homograph disambiguation accuracy in two ways: *micro-accuracy* and *macro-accuracy*. These two evaluation metrics are used to aggregate results across several groups. Here the groups in question are different homograph classes (e.g., *bass*, *live*, *record*, etc.). Micro- and macro-accuracy are equivalent when all groups are the same size. However, some homograph classes in the homograph data have more examples than other, so the two metrics will produce somewhat different results.

For both types of accuracy, we first compute lists of the number of correctly and incorrectly classified examples for each group. Alternatively, we compute the number of correctly classified examples for each group, and the number of examples (both correctly and incorrectly classified) for each group. Let us suppose that there are n groups, and denote the number of correctly classified examples by $C = c_1, c_2, \ldots, c_n$ and the size of the groups by $S = s_1, s_2, \ldots, s_n$. Micro-accuracy (or *micro-averaged accuracy*) is then the average accuracy across all examples:

$$\frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} s_i}.$$

Or in other words, micro-accuracy averages at the example level and largely ignores the structure of the different groups. In contrast, macro-accuracy (or *macro-averaged accuracy*) averages at the group level. It is given by the (arithmetic) mean of the individual group accuracies:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{c_i}{s_i}$$

Problem Using pencil and paper, or Python, compute micro-accuracy and macro-accuracy for the data set shown in Table 1.

C	12	19	40
S	21	41	53

Table 1: Counts of correct classifications C and overall sizes S for three groups.