

Assignment 4: audio event classification

Description:

Implement a classifier able to predict the audio event recorded in an audio excerpt.

Input:

As dataset use the **Urbansound 8k** which is available at the following links:

- <https://urbansounddataset.weebly.com/urbansound8k.html>
- <https://www.kaggle.com/chrisfilo/urbansound8k>

The dataset is divided into 10 **folds**, as the author suggests to work using **10 fold cross-validation** ([https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)#k-fold_cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#k-fold_cross-validation)).

In a nutshell, this means: select one fold for testing; use the remaining 9 for training; repeat this procedure 10 times (for all possible combination of testing folds).

The final score is obtained aggregating the scores obtained for each test split.

If you need to reduce computational complexity of the experiment, you can repeat the procedure 3 times rather than 10.

The results of the classification must be reported as a confusion matrix and, optionally, other metrics of your choice.

Output:

- a brief presentation of your work (max 5 minutes) that will be given to the class
- a more detailed report in which you illustrate and explain every step of your classification system and in which the results are shown and commented (max 8 pages) **to be delivered by May 14th**.
- a link to a repository containing the code (e.g. on GitHub) with minimal comments.