MLPC 2023 Task 3: Classification

Katharina Hoedt, Paul Primus, Florian Schmid, Jan Schlüter

Institute of Computational Perception
Johannes Kepler University Linz

April 17, 2023

1 Context

Our overall goal for this year's project is to train a system that can tell, for any instant in an audio recording, which bird species is audible (if any). Turning this into a machine learning problem entails:

1. splitting up the audio into small fragments, which will become the training examples,
2. computing a set of audio features for each fragment,
3. assigning a label to each fragment,
4. training a classifier to predict the label from the audio features.

Using your annotations from Task 1, we have prepared a dataset of 120,000 labeled audio fragments along with audio features. In Task 2, you explored the dataset, which included looking for correlations between features (to find features that are redundant) and correlations of features with labels (to find features that are useful). You will now use these insights to train classifiers.

2 Task Outline

Using the same training data as in Task 2, perform systematic classification experiments in your team:

- Focus on predicting the presence or absence of the 6 bird species. Remember that we made the simplifying assumption that these species will never be audible at the same time, so for each audio fragment, there are exactly 7 options: either one of the 6 species is audible, or none of them.

- Decide on an evaluation criterion to use. Mind that your goal is to correctly recognize all the birds.
- Apply at least least four different learning algorithms, from different major groups: Support Vector Machines, Neural Networks, Nearest Neighbor Classifiers, Naive Bayes, Decision Trees, Generalized Linear Models, Linear and Quadratic Discriminant Analysis... maybe also an "ensemble method" such as a Random Forest.
- For each algorithm, perform a systematic evaluation of different parameter settings, especially for those parameters that control the algorithm's overfitting behavior. Analyse and document how the parameters affect whether overfitting occurs (and to what extend it occurs), and how they affect classification performance.
- Use cross-validation for all your experiments. Decide how to split up the data into folds, based on what you know about the data. Mind that your goal is to estimate how your classifier will perform on unseen bird recordings.

For these experiments, you may use any machine learning package you want, such as scikit-learn for Python, WEKA for Java, or builtin toolboxes for Matlab. You may want to exploit that neither all the features nor all the training examples are equally useful. Especially when working on (not-so-powerful) laptops, it can pay off to subsample the training data to save computation time.

Document your results in a slide deck, touching each of the following aspects:

1. **Data split**: How was the cross-validation split done, and why?
2. **Features**: Which subset of features was selected or which preprocessing was applied, and why?
3. **Evaluation:** Which evaluation criterion did you choose to compare parameter settings and algorithms? What is the baseline performance, what is the best we can expect?
4. **Experiments:** For at least four different classifiers from the major groups given above, vary the parameters, focusing on: (a) Does overfitting occur, to what extent, and what does it depend on? (b) How does classification performance change?

Include visualizations and/or tables to present your results. Compile a slide deck of at most 20 slides (plus a title slide that includes your team name and member names). Make sure to address all four aspects in your report.

Submit your slide deck as a PDF on Moodle by **May 18th**. Only one team member needs to submit on behalf of the team.

3 Dataset

The dataset is the same as for Task 2. The dataset download links are available on Moodle, and the format and content of the dataset is described in detail in the slide deck for Meeting 2 (March 20). Please refer to that slide deck for information on the audio features, the derivation of the labels, and the file formats.

4 Grading

Reports for this task are evaluated according to the following criteria, for each of the five aspects given in the task outline:

- **Thoroughness and correctness:** Have you seriously thought about the problem? Are the proposed procedures and experiments sound, correct?
- **Presentation, Completeness, Clarity:** Are the ideas, features, algorithms, and results described clearly? Does the report contain all the information needed for the reader to reproduce the results (e.g., exactly which features were used, what were the parameter settings, ...)? Are results presented in a structured way (e.g., tables, graphics) that permits the reader to easily grasp them?
- **Punctuality:** The reports must be submitted in time. Any delay will result in reduced grades. Specifically, submitting on May 19 will deduct $1/3$ of the points, submitting on May 20 will deduct $2/3$ of the points, and submissions on May 21 or later will be rejected.
- **Additional bonus: Creativity:** Have you tried to explore alternative paths? Have you tried to come up with novel, creative ways of addressing the problem?

The data exploration task accounted for 20% of the grade; the classification task and challenge account for 40% each.

5 Task Deliverables

The only deliverable for this task is your team's report. Upload it as a PDF on Moodle by **May 18th**.