# Assignment 2 Introduction to Machine Learning

# Peer review

## Paper 1: Effectiveness of features and best classifiers for them

**Summary:**

The paper aims to identify features and classifiers that can predict specific music genres. Some literature is reviewed about how genre classification has been done by using acoustic features and lyrics. Several set of features, namely: lyric, metadata and audio features are tested separately with different set of classifiers. A Tag Naïve Bayes classifier, that calculate the probability of any given tag when a label is present, and Random Forest are used to test if lyric features (tags) can predict Punk and Folk genres, which both classifiers do according the author’s threshold of 50% f1 score. Later, metadata is analysed. First, loudness as predictor for metal genre hypothesis is tested using Gaussian Naïve Bayes and K-nearest neighbour, with, again, positive results. Then, other metadata is analysed without detailing. Finally, audio features are analysed using Multilayer Perceptron and Random Forest to, then, perform error analysis.

**Well done:**

Hypothesis presented are reasonable. The method is coherent with the hypothesis. Also, the method is systematic as different set of features are tested separately, and the use of every classifier is justified. Evaluation metrics are suitable for the task. Analysis of the classifiers evaluate their theorical strengths and weaknesses and classifier selection is supported. An ad hoc Naïve bayes classifier for tags is developed to work specifically with lyric features. Error analysis is introduced. Overall, the text is well structured and clear.

**Improvements:**

There is a lack of references in the pre-processing section about *minimum* and *max equalisation*. The pre-process could be explained better as is do not mention if the oversampling is done with replacement or not. Why the author chooses Zero-R as baseline is not mentioned either and then is never mentioned again. Explanation of Tag Naïve Bayes may need re-wording. Other metadata features tests are not described enough, details of which classifiers were tried or if any other transformations were made are missed. Analysis of one-dimension k-NN may miss an important factor such as data distribution and outliers. Hypothesis about Multilayer Perceptron and Random Forest performing ‘well’ over audio features is vague. There are some misspellings.

## Paper 2: A Report on Machine Learning for Music Genre Classification

**Summary:**

The paper problematizes over the issue of music genre classification as a subjective task. It sets as main objective the problem of classifier algorithms efficiency and how feature selection influence those classifiers. K-nearest neighbours is presented and described, including advantages and disadvantages related to the algorithm efficiency and expected performance based on literature. Then, Decision tree is presented, describing how works including the formula of *Information Gain* for feature selection.

In the next section a methodological plan for experiments is described as follows: The dataset is described by its attributes and labels and divided in three sets, metadata, lyric and audio features. Then, data pre-processing is described. Tags are ‘sort-of’ one-hot transformed, numerical are discretized and labels are changed to numerical. Later, the selected algorithms are executed using threefold cross validation and hyper-parameterization, testing several combinations of feature set. Also, Zero-R baseline is selected. Algorithm performance is analysed and some conclusion over the effect of different set of features discussed.

**Well done:**

There author executes a systematic test of different combination features using a comparative baseline and a set of algorithms. Textual data is pre-processed.

**Improvements:**

No theorical evaluation of decision tree is provided. In general, no reflexion or critical analysis of the algorithms over the dataset. No justification over baseline, evaluation metrics or classifiers is provided. No discussion over classes or its distribution. No attempt to deal with the objective of analyse algorithm efficiency. No empirical support for some analysis, on the contrary, some claims contradict the results.

The format and language do not follow the instructions and do not comply with academic writing guidelines. The document contains many errors and incongruences, such as grammar mistakes and changing font sizes, font types and bibliography reference systems. Wording needs improvement as some parts are difficult to understand, particularly pre-processing subsection. There is no error analysis.