

INFO284 Machine Learning Exam, spring 2022

Delivery date: May 13th 2022, 14:00

Format: Jupyter notebook (ipynb-file) containing runnable Python code, documentation and reflections on process and result.

Word limits: The total text parts should not be more than 3000 words. There are no limits on Python code size.

Machine learning on music album data

At <https://filesender.uninett.no/?s=download&token=80a1600f-c788-4fda-b25a-cf1cec10eae4> you will find a tabular data set that contains information about 2833 music albums. These data are a subset of a data set found at kaggle (<https://www.kaggle.com/elemento/music-albums-popularity-prediction>). In addition, you will find a data set that contains the album cover for those 2833 albums.

- a) You are supposed to **build at least four machine learning models from the tabular data** to predict popularity of the albums based on the features in the tabular album information data set. You may take inspiration from the example code at kaggle.com or other web sites when building models for these data, but **if you do you need to refer to these examples**, and you need to explain how you used and extended these approaches in your own solution.
- b) You shall also **build a (convolutional) neural network model from the 2833 images to predict a selection (of own choice) of three features and the popularity score from the tabular data set**. To do this you are recommended to use a **pre-trained convolutional neural network** (of your own choice), but if you have available computational power, you may of course try to build your own complete CNN.

Some links to information about pre-trained CNN:

- <https://towardsdatascience.com/4-pre-trained-cnn-models-to-use-for-computer-vision-with-transfer-learning-885cb1b2dfc>
- <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>
- <https://medium.com/@mikhailenko/instructions-for-transfer-learning-with-pre-trained-cnns-203ddaefc01>
- BOOK: F. Chollet. Deep Learning with Python. Ch. 5.3 Using a pretrained convnet. p.143-159

You shall deliver code in the form of a **well commented Jupyter notebook**. This code needs to run on the original data set, so any preprocessing you choose to do needs to be programmed in Python and included in the notebook. The code shall in the end return the results of your experiments with your chosen models. You need to explain

- Important and relevant properties of the data
- how you preprocessed data like which features you selected, did you do dimension reduction, how you reformatted data, etc.
- how you decided on parameters for your machine learning models, did you use any regularisation techniques.
- how the methods were measured and compared

Please inform about any special Python libraries that need to be installed to make your code runnable.

Finally, as a concluding comment in the Jupyter notebook, you need to write a summary of your results, and discuss consequences of such results.

It is not necessarily so that high scores for machine learning models will give a good grade on your report, or vice versa, low scores a bad grade. What counts, is a well-argued, well described and smart machine learning investigation from start to end. The problem may in fact be so that it is not possible to get really good results on these data.

Final note: These data are prepared for this course, and are shared with you in confidence that you do not share them in any way, but use them only for the purpose of this exam.