STAT40970 – Machine Learning & A.I. (Online)

Assignment 2

Deadline - Wednesday 26 April at 17:00

Indoor scene recognition

The folder `data_indoor.zip` contains images concerning different indoor scenes from rooms and locations commonly present in a standard family home. The task is to predict the type of room/scene present in the image. Indoor scene recognition is a challenging problem since some indoor scenes can be well defined by global spatial and structural properties, while others are better characterized by the objects included in the space. The dataset is a subset of a larger dataset for indoor scene recognition. More information is available here: `http://web.mit.edu/torralba/www/indoor.html`.

The images are divided into `train`, `validation`, and `test` folders, each containing the folders related to the type of the room in the image (i.e. the categories of the target variable): `bathroom`, `bedroom`, `children_room`, `closet`, `corridor`, `dining_room`, `garage`, `kitchen`, `living_room`, `stairs`. The number of images available for each scene is variable and it ranges from 52 to 367 in the training set, with validation and test sets being roughly half the size.

Task

The task is to build a predictive model to predict the type of indoor scene from the image data.

1. Deploy at least 4 different deep learning systems characterized by different configurations, hyperparameters, and training settings (architecture, number of hidden units, regularization, kernel size, filter size, optimization, etc.). These deep learning systems can be of the same type, for example 4 different DNNs characterized by different architectures and settings, or of different types, for example 2 DNNs and 2 CNNs with different settings. Motivate clearly the choices made in relation to the settings, configurations, and hyperparameteres used to define the different deep learning systems. *(30 marks)*
2. Compare appropriately the deep learning systems considered, evaluating and discussing their relative merits. Comment on their training and predictive performance, and select the best model a predicting the type of indoor scene from the data. *(60 marks)*
3. Use the test data to evaluate the predictive performance of the best model. Comment on the ability of the model at recognizing the different scenes. *(10 marks)*

Instructions and guidelines

- **Discuss and motivate the various decisions taken in all stages of the model building process**.
- Encode the images in numerical RGB tensors of width/height of 64×64.
- Not all data augmentation transformations make sense for all scenes, and some particular transformations might be more appropriate.
- You could use `tfruns` to specify and tune the different models, but it would require **a lot of computing time**. No need to do this for full marks. You can take a trial and error approach to tune the systems, but you must motivate the various decisions taken.
- If you want, you can consider aggregation of some similar scene classes. Also in this case you **must clearly motivate your choice** and why some classes are aggregated.

- You will not be evaluated on the basis the predictive performance of your models, but you would need to show that attempts have been considered to build a system with reasonable performance.
- Submitting only code will not give any marks.

Submission rules

- Write a short report and submit it as a single pdf file (approximately max 8-10 pages, code excluded).
- Include the R code used for analysis in the report. The report can be produced using R Markdown, with the code included in the main text or as an appendix. **The code must be working and the analysis must be reproducible in all parts**.
- Multiple submissions before deadline are allowed and only the latest one will be considered for marking.
- Submission after deadline will incur in penalization as UCD rules (see "Module details" document).
- **Plagiarism is strictly prohibited** (see "Module details" document and "Information materials" tab).