UNIVERSITY OF SOUTHERN DENMARK
FACULTY OF SCIENCE

DS807: Applied Machine Learning

Winter 24/25 (ordinary exam)

1. Assigner: Christian Møller Dahl.
2. Hand-out: 16th December, 2024, 12:00 (noon).
3. Hand-in: 31st January, 2025, 12:00 (noon).
4. All pages, incl. the front page, should contain the following: Full name and date of birth (**not** CPR-number).
5. All pages must be numbered.

Form of examination for the certificate:
Take-home assignment.

Supplementary information for the form of the exam:
The exam may be solved in groups of up to 5 students or individually. Working in groups is encouraged. You should make the definition of your group in System DE-Digital Exam. Follow this guideline.

Further:

1. In your report, state explicitly who is responsible for which parts.
2. Location: Home assignment.
3. Internet access: Necessary.
4. Hand-out: System DE-Digital Exam.
5. Hand-in: System DE-Digital Exam.
6. Extent: No longer than 30 pages, excluding references, appendices, and code.
7. Exam aids: All exam aids are allowed.
8. File format: The report must be submitted as a **.pdf** file. The code may be submitted as one of: **.ipynb**, **.html**, or **.py**.

Grading according to the Danish 7-point scale. Grading based on the performance of the individual student compared to the learning goals.

Exam questions

For all problems in the exam, ensure you explain how and why you prepare the data. This includes detailing considerations for further splitting, scaling, reshaping, and other preprocessing steps. In particular, it may be necessary to limit data and models due to hardware limitations on your system. This is acceptable, but you should discuss the potential implications and consequences of these limitations. Generally, if the same method is used for multiple questions, it is sufficient to describe the procedure once and refer to it in subsequent questions; however, the rationale for using the method must still be provided.

Problem 1:

During the semester, you have become deeply intrigued by the field of dermatological imaging, a vital domain within medical AI. This area is evolving rapidly thanks to advancements in deep learning and publicly available datasets. You recently studied the paper by Codella et al. (2018) and were fascinated by the International Skin Imaging Collaboration (ISIC) dataset, which has been foundational for automated melanoma detection. This dataset has been curated to benchmark algorithms for tasks such as lesion segmentation, dermoscopic feature detection, and disease classification.

The primary objective of this exam is to design, train, and evaluate models for a general skin disease classification using the ISIC dataset. The dataset includes images labeled with 14 categories: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, chickenpox, cowpox, dermatofibroma, hfmd, healthy, measles, melanocytic nevi, melanoma, monkeypox, squamous cell carcinoma, vascular lesions. Training, validation, and test splits are provided to support your modeling efforts. The training set includes 29,322 images, the validation set includes 3,660 images, and the test set includes 3,674 images. You must use the training/validation images to train/validate models, while the test images must be used only to evaluate performance.

The ISIC dataset is available from Hugging Face. Further instructions and hints on how to load the data efficiently and flexibly are available on the course site's itslearning platform.

<u>Question 1:</u>
Use shallow learning to perform skin lesion classification. Specifically, you must:

1. Briefly discuss how the skin lesion classification problem can be addressed using classical shallow machine learning techniques, such as support vector machines, random forests, and gradient boosting.
2. Implement one of these methods, as well as a combination of them, for the classification task. Present and discuss relevant performance measures. Evaluate the performance of the models and discuss performance differences, including why performance might vary across skin classes and across models.
3. Consider and discuss the role of image resolution in your models. Does higher resolution significantly impact performance across skin lesion classes and models?

<u>Question 2:</u>
Use deep learning to perform skin lesion classification. Specifically, you must:

1. Briefly explain why convolutional neural networks (CNNs) are well-suited for skin lesion classification based on images?
2. Train one or more CNN architectures to classify the skin lesions. Specifically:
    a. Discuss and implement alternative CNN architectures. Get inspired, for example, by the ResNet and EfficientNet architectures and the various “blocks” they are building on.
    b. Experiment with different optimizers (e.g., SGD, Adam) and visualize their performance. Motivate your final choice.
    c. Compare the impact of regularization methods like dropout, weight decay, and early stopping on training and validation losses. Discuss regularization and its relation to overfitting and provide visualizations.
    d. Explore data augmentation techniques and their effect on overfitting. Visualize the training process with and without augmentation.
    e. Apply transfer learning, detailing your approach and any required adjustments to the data. Here, be sure to visualize plots of train and validation losses and accuracies.
    f. Evaluate the importance of image resolution relative to model depth and width, referencing EfficientNet principles (Tan and Le, 2020, EfficientNet).
3. Select your preferred CNN model(s) and justify your choice. Evaluate its performance on the test data using the same performance measures as for the shallow machine learning models in Question 1.

Question 3:
To solve this problem, focus on your preferred CNN model from Question 2. The objective is to better understand the model and investigate whether it can assist dermatologists in identifying selected skin lesion classes of your choice (for example, melanoma). Specifically, you must:

1. Visualize activation maps (e.g., Grad-CAM) for your preferred CNN model and for your selected skin lesion class.
2. Based on your potentially limited domain knowledge and your role as a data scientist, discuss whether these maps help identify regions of interest corresponding to your selected skin lesion classes (e.g., melanoma). Evaluate their potential utility for dermatologists and whether a CNN model could assist clinicians in identifying skin lesion classes in an effective and trustworthy manner.

Problem 2:

In the field of Natural Language Processing (NLP), emotion recognition is a crucial task with applications in sentiment analysis, mental health monitoring, and human-computer interaction. Building robust models for emotion classification requires datasets that capture diverse linguistic expressions and context-specific nuances.

You recently studied the paper by Saravia et al. (2018), which introduced the CARER framework for contextualized affect representations in emotion recognition. Inspired by this, you aim to apply emotion recognition techniques to analyze posts from the Bluesky social media platform, which presents a unique mix of personal and public discourse.

To achieve this, you will (through a series of questions and suggested tasks):

- Train models using the DAIR AI Emotion Dataset, which includes textual data labeled with 6 basic emotions: joy, sadness, anger, fear, love and surprise.
- Use your trained model(s) to classify emotions in a sample of posts from this Bluesky Dataset, exploring how well your model(s) generalizes to new and unseen data.

Questions and suggested tasks:

1. Data Exploration and Preprocessing:
   Question: What patterns and distributions of emotions are present in the DAIR AI Emotion Dataset? How will you preprocess this data to prepare it for training emotion recognition models?
   Suggested Tasks: Conduct an exploratory data analysis (EDA) to justify the training and validation datasets provided. Is the train-validation split appropriate? Include visualizations showing the distribution of emotions, word frequencies, and text lengths. If needed change the training-validation split and discuss and apply preprocessing steps. Discuss any important patterns or insights you observe from the EDA.

2. RNN Model for Emotion Recognition:
   Question: For the emotion recognition task, you are now required to develop and train models using Recurrent Neural Networks (RNNs). Why are RNNs suitable for this task, and how do they perform?
   Suggested Tasks:
   a. Model Development: Describe the architecture of your RNN model(s). Include details such as the type of RNN cells used (e.g., LSTM, GRU), the number of layers, and any other relevant architectural choices.
   b. Data Preparation: Explain how you prepared the textual data specifically for the RNN model(s). If any, discuss additional preprocessing steps or transformations that were necessary to make the data suitable for RNN training. Include details on vocabulary size, sequence length and embedding dimensions.
   c. Training Process: Outline the process you followed to train the RNN model(s). Include details on the training parameters such as batch size, number of epochs, choice of optimizer, the presence of dropout and other regularization techniques. As in **Problem 1** visualize how all these factors affect training and validation losses and accuracies.
   d. Performance Evaluation: Evaluate the performance of your RNN model(s) on both the training, validation, and test data. Present key metrics such as accuracy, precision, recall.
   e. Insights and Analysis: Provide an analysis of the RNN model's performance. Discuss any challenges you faced while training the RNN and how they were addressed. Reflect on the suitability of RNNs for the emotion recognition task based on your results.

3. Applying the model to Bluesky posts:
   Question: Apply your preferred trained RNN model(s) to classify emotions in a sample of Bluesky posts. How well does your RNN model(s) generalize to this dataset, and what challenges arise when analyzing these posts?

Suggested Tasks:

a. Preprocess a selected sample of the Bluesky dataset to align with the format used for training. Motivate your choice of sampled Bluesky posts.
b. Use the trained model to predict emotions from the Bluesky posts. Compare the predicted emotion distributions with those in the DAIR AI Emotion Dataset. Present and discuss your main findings.
c. Discuss challenges such as linguistic variability, domain shift, and the need for further fine-tuning and provide recommendations on whether your RNN model is appropriate for emotion recognition in Bluesky posts.

4. Insights and Improvements:
   Question: Reflect on the results and discuss how the emotion recognition model could be improved for future applications. Are there any findings you want to highlight?
   Suggested Tasks:
   a. Identify limitations of the current model, such as overfitting, underfitting, or inability to capture context-specific nuances.
   b. Propose strategies for improvement, such as domain adaptation, use of pre-trained language models, or collecting additional labeled data.

Loading data (important!)

As already mentioned above I will provide hints/instructions on how to download and prepare the data. I have posted these on the course site's itslearning platform under the plan named: "**Exam: Instructions/hints on how to load data**". It is important that you look at the notebooks.