

21.11.24

Machine Learning With TensorFlow

CONVOLUTIONAL NEURAL NETWORKS IN TENSORFLOW, PART 1

- Quiz
- Assignment
- Literature Review
- Open Discussion
- Input
- Project Milestones
- Tasks for next week

QUIZ

Quiz



<https://forms.office.com/e/bwec7rtVi2>

ASSIGNMENT

ASSIGNMENT: WHO WILL PRESENT NEXT?

Literature Review - Breakout

- Shortly present your project idea and the papers, articles, tutorials you found to the other group(s)
- Discuss interesting ideas and concepts of these resources
- Discuss the process you have taken while searching for these resources (e.g. usage of search engines, chatbots, etc.)

Open Discussion

- Beside image data augmentation through transformations of the training data presented in the videos, what other forms of extending the diversity of an existing image dataset can you think of?

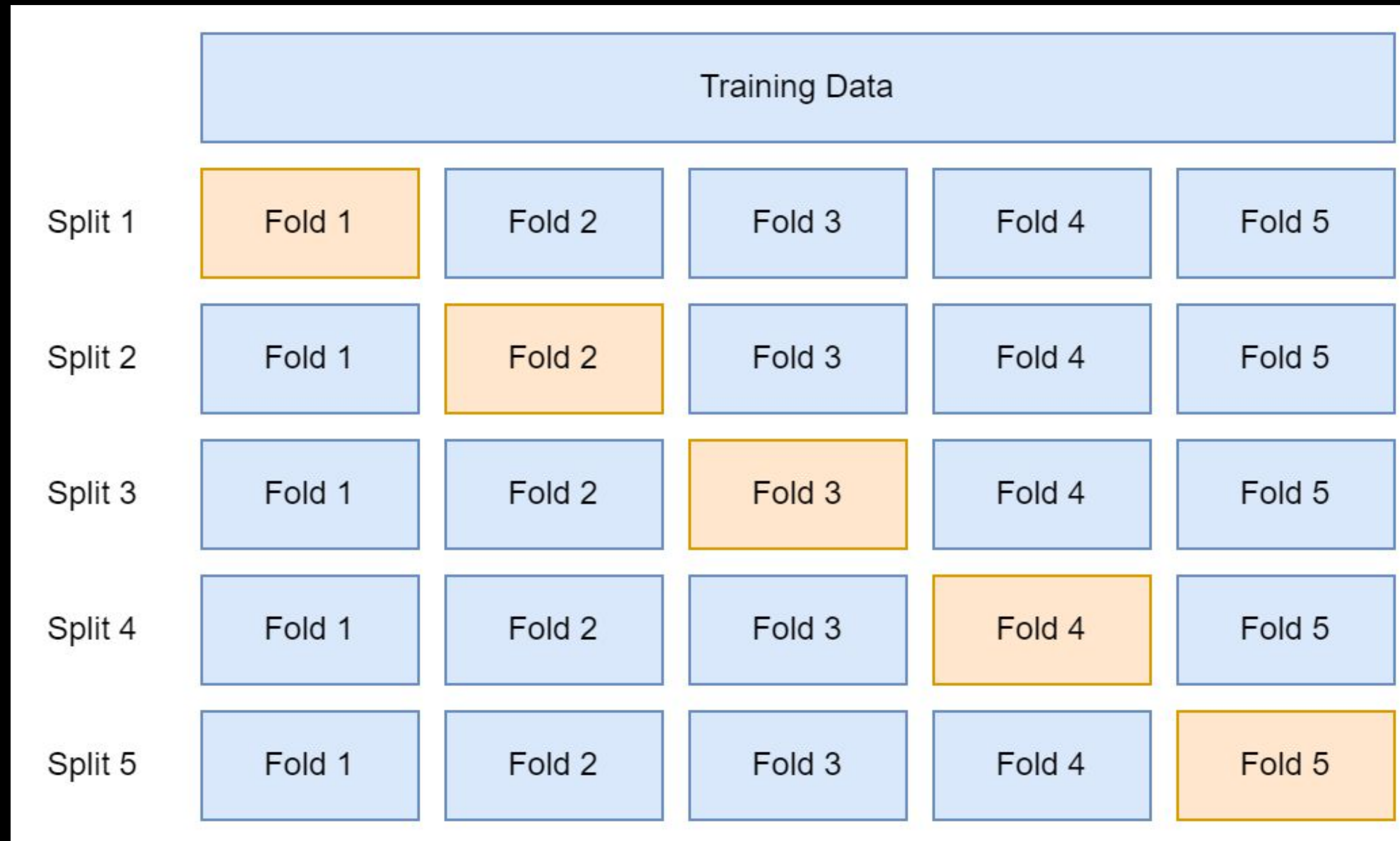
Open Discussion

- Beside image data augmentation through transformations of the training data presented in the videos, what other forms of extending the diversity of an existing image dataset can you think of?
- What are the main differences between extending a dataset through (synthetically) generated images and performing image data augmentation during the training process?

Input - Cross-Validation

- Hold-out dataset: Split between training and validation data (and sometimes test data) → Time Series
- How do we know if the initial split is the best possible split? → We don't, therefore we use cross-validation
- k-Fold cross-validation:
 - Split training data into multiple folds
 - Train and test on different training/validation data combinations (splits)

Input - Cross-Validation



Input - Cross-Validation

- Stratified k-Fold cross-validation: Creates folds that keep the label/target distribution → e.g. for skewed classification datasets
- Group k-Fold cross-validation:
 - Special case where multiple samples belong to the same group (e.g. images of a patient)
 - Possible data leakage if samples of the same person are in training and validation data
 - → Group k-Fold cross-validation

Further information: <https://github.com/abhishekrthakur/approachingalmost/blob/master/AAAMLP.pdf>

Project Milestones

- 07.11. Project Pitches
- 14.11. Form Groups
- 21.11. Literature Review
- 28.11. Dataset Characteristics
- 05.12. Baseline Model
- 12.12. Model Evaluation
- 19.12. (Feedback Session)
- 02.01. Final Model Optimizations
- 09.01. Final Model Optimizations
- 16.01. Project Presentations
- 23.01. Project Presentations

Project Milestones

- 07.11. Project Pitches
- 14.11. Form Groups
- 21.11. Literature Review
- 28.11. Dataset Characteristics
- 05.12. Baseline Model
- 12.12. Model Evaluation
- 19.12. (Feedback Session)
- 02.01. Final Model Optimizations
- 09.01. Final Model Optimizations
- 16.01. Project Presentations
- 23.01. Project Presentations

Project Milestones

- 07.11. Project Pitches
- 14.11. Form Groups
- 21.11. Literature Review
- 28.11. Dataset Characteristics ←
- 05.12. Baseline Model
- 12.12. Model Evaluation
- 19.12. (Feedback Session)
- 02.01. Final Model Optimizations
- 09.01. Final Model Optimizations
- 16.01. Project Presentations
- 23.01. Project Presentations

Next Milestone: Dataset Characteristics

- Perform a data analysis of your dataset to identify:
 - Key characteristics of the dataset (Original data sources, number of samples, etc.)
 - Potential biases arising from the collected data
 - Outliers in the dataset
 - Missing values
 - Feature distributions
 - Correlations
- Utilize graphics, plots and tables to summarize the main characteristics of the dataset.

TASKS UNTIL NEXT WEEK

- Complete week 3 and 4 of the course "Convolutional Neural Networks in TensorFlow"
- Complete the assignment notebook
- Analyze the characteristics your dataset and add the analysis to your GitHub repository