Deep Learning - Assignment 3

Group 2

Instructor:
Siamak Mehrkanoon

Group Members:
Sheng Kuang (i6237193)
Yimin Yang (i6246099)

1. Navigation to Answers

(a): Which model do we use?

A1: CNN + Softmax.

(b): How is the comparison of the accuracy of the 2 types of classification?

A2: (Table 3)

(c): What additional approaches do we select?

A4: Transfer learning to improve accuracy of cross-classification.

(d): What are the key hyper-parameters (and interpretations)?

A3: Latent Channel Size, Dropout, L-1 Penalty (Table 2, interpretations in Section 4.3)

2. Motivation of Model Selection

2.1 CNN + Softmax

The provided MEG data consists of signals of 248 channels. These channels located in different spatial positions represent complex brain activities at a certain time point. According to the model proposed by Daunizeau et al. in 2007, the latent space of MEG signals could be generated by a linear transform [1]. To catch the event-related potentials(ERP) of four human activities, typical CNN architectures(i.e., a 1-D convolutional layer and a max-pooling layer) are adopted to extract the temporal and spatial features from each latent channel [2]. Finally, a fully connected layer followed by a softmax layer is applied to classify the status.

2.2 Transfer Learning

Although the above model performs well in intra-classification, the cross-classification test received unsatisfying precision. In this case, we propose the model with the same architecture which is trained by the cross-training dataset as a common feature selector, then use the intra-training dataset to fine-tune the last fully connected layer to get a more accurate classification result. Moreover, a dropout setting is added to the fully connected layer and L-1 regularization is applied to make the model more robust.

3. Method

3.1 Pre-processing

3.1.1 Hypothesis Each channel of MEG signals has dependencies and could be transformed into latent MSG source by linear transformation[1].

3.1.2 Normalization The MEG data are downsampled to 127.125 (=2034/16) Hz in order to reduce calculations but not reduce the performance significantly. The normalization approach we selected is ***Z-score Scaler***. Each channel of subjects is normalized independently after downsampling.

3.1.3 Training & Validation Dataset The intra-classification training dataset consists of 32 trials from a single subject. The cross-classification training dataset consists of 64 trials from two subjects. No validation dataset is applied to both training processes.

3.2 Architectures

The CNN+Softmax architecture we use is listed in Table 1. The loss function we choose is the ***CrossEntropyLoss*** while the ***L-1*** regularization weight is initialized to 0.0001. The initial batch size is 8 due to the training size.

Table 1. The architecture of the proposed model in intra-classification.

Layer Name	Input Size	Output Size	Settings & Parameters
Latent Source Layer	Batch×2226×248	Batch×2226×32	Linear transform
1-D Convolutional Layer	Batch×2226×32	Batch×2226×32	Kernel, Stride, Pooling, Padding, ReLU after convolution
MaxPooling Layer	Batch×2226×32	Batch×139×32	AdaptiveMaxPooling
Fully Connected + Softmax Layer	Batch×139×32	Batch×4	Dropout

3.3 Experiment Setting Up

The best hyper-parameter combination of each model is based on the fine-tuning experiments. For all the experiments, the learning rate (LR) is set to 0.0001 to minimize the cross entropy loss. In addition to LR, the following hyper-parameters and settings are investigated:

- Kerner Size: 3, 5, 7, 9, 11
- Latent Channel size: 4, 8, 16, 32, 64
- Stride: 1, 2, 3, 4
- Pooling Factor: 2, 4, 8, 16, 32, 64
- Dropout: 0.25, 0.5, 0.75, 0.9
- L-1 Regularization: 0.01, 0.001, 0.0001

The convergence speed of the intra-model is shown in Appendix A. The best hyper-parameters are presented in Table 2.

Table 2. The best hyper-parameters of two scenarios.

Hyper-parameter	Intra-classification	Cross-classification
Latent Channel	32	32
Kernel Size	7	7
Stride	1	1
Padding	2	2
Pooling Factor	16	16
Batch Size	8	8
Dropout	0.25	0.9
L1 weight	0.001	0.0002
Learning Rate	0.0001	0.0001

4. Results

4.1 Intra-classification

The accuracy of the naive intra-classification test is 62.50%, which means 5 out of 8 trials are classified correctly. The intra-model without the dropout setting and L-1 regularization (normal) has the same accuracy as the other one (i.e., dropout). Meanwhile, the model with the dropout probability equals 0.25 has the lowest cross-entropy loss. However, two models with L1 penalty reach 87.5% accuracy, which is the most satisfying result among intra-classification. The interpretation of this phenomenon is that the output of the max-pooling layer could still be regarded as a sparse MEG feature space. The white noise and interference are suppressed and eliminated by adding the L1 penalty to the loss function. The aforementioned results are listed in Table 3 along with cross-classification results.

4.2 Cross-classification

If we apply the same model to the cross-classification problem, the test accuracies are relatively low compared to the intra-classification problem. The highest accuracy reaches 47.92% when predicting the statues among multi subjects. The solution to tackle this obstacle is to train the model in two steps. The first step is to pre-train the model by using trials from multiple subjects. The second step is to fix weights of all layers except the last fully connected layer, and then fine-tune the layer by using the data of the intra-classification task. These steps aim to force the last fully connected layer to focus on similar ERG features rather than external interference.
According to the result, the cross-classification model with dropout setting and L1-penalty reaches the best accuracy at 62.50% in the pre-training step. The model with the dropout probability of 0.9 outperforms other models after fine-tuning the last fully connected layer. The best classification accuracy in experiments reaches 70.83%.

Table 3. The best performance of intra-classification after 500 epochs and cross-classification models after 1000 epochs (each model is tested for 6 times).

Scenario	Model Setting	Loss	Accuracy(%)
Intra-classification	Normal	0.00040229	62.50
	Dropout	0.00017145	62.50
	L-1 penalty	0.24968345	**87.50**
	Dropout + L-1 penalty	0.25206053	**87.50**
Cross-classification	Normal	0.00001594	47.92
	Dropout	0.00357825	60.42
	L-1 penalty	0.18251026	45.83
	Dropout + L-1 penalty	0.12429191	**62.50**
Cross-classification with transfer learning	Normal	0.00025362	50.00
	Dropout	0.00722692	**70.83**
	L-1 penalty	0.19723721	56.25
	Dropout + L-1 penalty	0.13118136	68.75

4.3 Interpretation of key hyper-parameters

Dropout & L1 Regularization The dropout probability in the intra-classification (0.25) is much less than in cross-classification (0.9) because the individual differences are more significant between subjects. The MEG signal of different people may have different frequencies and evoked oscillations. Meanwhile, the data collection process might be disturbed by the environment or individual little actions (eye movement). So the dropout and L-1 regularization could suppress the interference and eliminate the overfitting problem.

Latent Channel The latent channel of MEG source could be regarded as the signal restoration from the leads to brain activities. 32 channels represent principal components of the MEG.

5. Conclusion

In this assignment, we explored the CNN model together with the dropout and L1 regularization techniques in classifying the subject states by using MEG data. The best intra-classification architecture we investigated consists of a linear layer to transfer the latent MEG source, a 1-D convolutional layer to extract temporal features, a max-pooling layer to combine spatial information, a fully connected layer with dropout, and a softmax layer to project features into four classes. The accuracy of the intra-classification model with L-1 regularization reaches 87.50%. As for the cross-classification model, a transfer learning method is adopted to fine-tune its last fully connected layer. The dropout is set to 0.9 to suppress the overfitting problem. The model only with the dropout setting reaches the best accuracy of 70.83% from the experiment.

6. Reference

[1] Daunizeau, Jean, and Karl J. Friston. "A mesostate-space model for EEG and MEG." NeuroImage 38.1 (2007): 67-81.

[2] Zubarev, Ivan, et al. "Adaptive neural network classifier for decoding MEG signals." Neuroimage 197 (2019): 425-434.

Appendix

A. Intra-model convergence speed with different hyperparameters