problem in EEG scenarios. To alleviate this problem, we embed Dropout into the proposed models. Specifically, in an epoch during the training process, Dropout randomly selects kernels and sets the output to zero according to the dropout rate in each layer [106]. This method ensures the networks do not overly enhance the connections that fit non-generalizable noise and only focus on generalizable EEG patterns. During testing, dropout is a transparent module, and the unbiased networks guided by dropout are used to evaluate the test performance. Commonly, dropout modules are deployed after nonlinear activations.

Exponential Rectified Linear Unit

Exponential Rectified Linear Unit (ELU), which is a commonly-used nonlinear activation function, is described as eq.(2.20),

$$f(x) = \begin{cases} x & if\ x > 0 \\ \lambda(e^x - 1) & otherwise \end{cases} \tag{2.20}$$

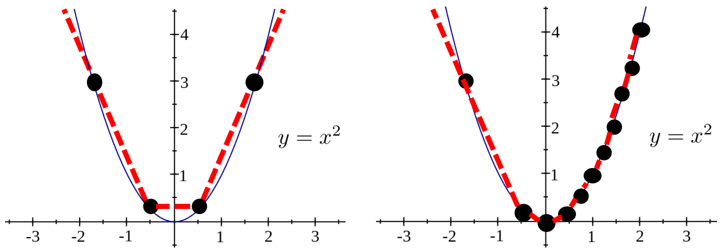
where $\lambda(\geq 0)$ is hyper-parameter. Instead of setting the output for a negative input to zero as ReLU [76], ELU provides a smooth exponential decrease. According to [56][92], ELU, which is an alternative to ReLU, achieves better performance in EEG-related tasks.

2.3.5 Model evaluation

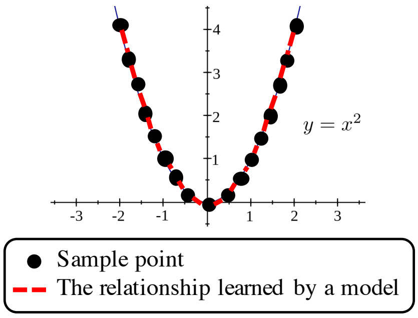
The requirement of data

During training, a deep learning model minimizes the difference between the predictions and the labels. By using Gradient Descent to adjust weights, the prediction ability gradually approaches the optimal. This optimization process is a learning process of a data-to-label mapping. Mathematically, a mapping relationship can be understood as a function, and the data and labels are the inputs and outputs of the function, respectively. Therefore, model optimization is a learning process of a function.

From the perspective of data, the performance of the function learning depends on the sample volume and distribution. For example, assuming data point x in the dataset D and the corresponding label y has a function relationship of $y = x^2$. If the data collection volume is small, then the function relationship is under-determined. Therefore, the model can only roughly learn the mapping, as shown in Figure 2.18(a). If the data collection volume is large but with imbalanced distribution, then the function is not fully presented, which leads to an incomplete mapping, as shown in Figure 2.18(b). Therefore, to fully reflect the relationship between data and labels, the data collection should be large enough with a reasonable sampling distribution, as shown in Figure 2.18(c).



(a) Sampling volume is small. (b) Sampling volume is large but with imbalanced distribution.



(c) Sampling volume is large with balanced distribution..

Figure 2.18: The learning of a data-to-label mapping from the perspective of data. If the sampling volume is small, the learning is coarse; if the sampling volume is acceptable but with an imbalanced distribution, the learning is incomplete; only when the sample volume is large with a reasonable sampling rate, the learning is successful.

Training, validation, and testing set

The evaluation process of a model contains training, validation, and testing. For a data set that meets the requirement above, we first divide it into training and test sets. The test set, which is not involved in training and validation, is only used for the performance evaluation. For the training set, traditionally, all the data are used for training to achieve good performance. Then, the test set is used for evaluation. However, the model may over-focus on the details in the training set, which results in the ignorance of general features in the data and poor performance on the test set, i.e., the overfitting problem. Therefore, how to monitor training statuses and prevent overfitting? Also, to make sure good performance on both training and test sets, how to pre-test the performance of the model and tuning parameters accordingly before testing?

The solution is to divide the training set into training and validation sets. Similarly, the training set is used for model training. Meanwhile, the validation set is taken as a simulated test set. For each training epoch, the performance on the training and validation sets are illustrated to monitor the model optimization, adjust parameters, and prevent overfitting accordingly. Moreover, the validation set provides a test simulation. We can tune the model on both training and validation sets and then deploy the model in the test environment.

The key to the allocation of the training, validation, and test sets is a consistent feature distribution. Inconsistent distributions may lead to inaccurate model evaluation. For example, when a dataset is relatively small, the amount of data that contains different features is relatively small, and the possibility of uneven distribution increases. That means a model that performs optimally under a training-validation-test allocation may not perform optimally under another allocation. Therefore, only models that perform well under multiple assignments can be considered as stable models with excellent statistical performance. The method of multiple sets of data allocation on small datasets is called cross-validation.

Cross validation

Practically, data collection may be subject to many factors, e.g., a complex collection process and a long acquisition cycle, which results in small data sets. For example, EEG data acquisition requires a large amount of prerequisite preparation, e.g., recruiting subjects, equipment adjustment before collection, and participants' practice in tasks. Therefore, the size of the EEG datasets is usually small. The scale of three EEG datasets are shown in Table 2.3.

Dataset	Number of participants	Number of samples
Conflict-containing EEG	81	about 2400
EEG Motor Imagery Dataset in PhysioNet	108	about 3500
BCI Competition IV Datasets(2b)	9	about 2100

Table 2.3: The scale of EEG datasets.

When the scale of a dataset is relatively small, cross-validation is required instead of a simple training-validation-test allocation described above. Firstly, the dataset is divided into training and test sets. Similarly, the test set is only used for model evaluation. Then, the training set is divided into several groups. In each fold, according to an allocation order and a training-validation ratio, several groups are combined to form a training set, and the rest forms a validation set. A sample in a training set may become a sample in the validation set in different folds, which makes a 'cross' validation. A data allocation with a train:valid:test ratio equals 4:1:1 is shown in Figure 2.19.

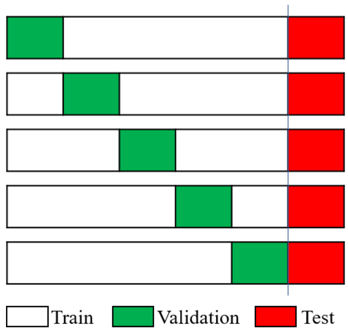


Figure 2.19: A 5-fold cross-validation scheme with a train:valid:test ratio equals 4:1:1.

2.4 EEG analysis using machine learning or deep learning

EEG is a representation of brain activities. By analyzing EEG data, we can recognize and analyze emotion, motor imagination, and pathological features in human brains. Since machine learning and deep learning methods have strong data mining ability, this section will introduce EEG analysis using machine learning and deep learning regarding the main research directions.

2.4.1 EEG-based emotion recognition

EEG has strong correlations with emotions in human brains [42]. The most commonly used EEG-based emotion dataset is DEAP [53]. Therefore, this literature review regarding emotion recognition mainly based on this dataset. Although we did not use the same dataset, we believe the feature extraction methods and the machine learning models are useful references for this thesis.