Building an EEG-based Alzheimer's Diagnostic: What is the Optimal Cross-Validation Method?

Colin Quirk[1], Geoffrey Brookshire[1], Yunan "Charles" Wu[1], and Ché Lucero[1]
[1]SPARK Neuro Inc., New York, NY, USA

SPARK NEURO

ALZHEIMER'S ASSOCIATION AAIC 22 ALZHEIMER'S ASSOCIATION INTERNATIONAL CONFERENCE®

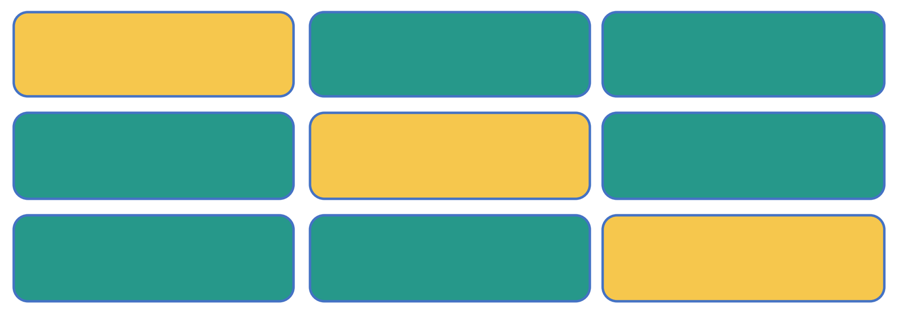
Background

Machine learning is a powerful tool for generating predictions about AD

When creating machine learning models, it is important to avoid overfitting to ensure good performance in a clinical setting

Here we explore the optimal method to avoid overfitting in simulated data

What is Cross-Validation?



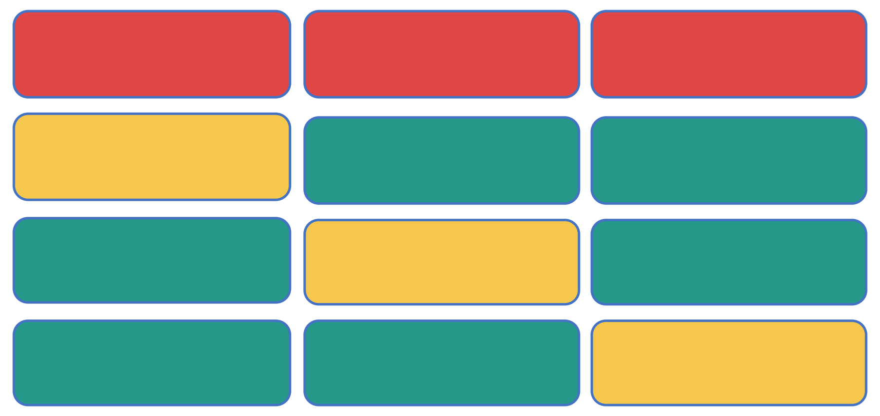
Full CV Loop

Individual chunks of the data are held out one at a time, so that all data is eventually treated as train and test

Model results are then averaged over all of the "folds" of the data

Nested Cross-Validation

While cross validation is a useful tool, many rounds of model fitting can still lead to overfitting and poor results



Single NCV Outer Loop Iteration

In nested cross validation, an additional outer loop is performed to avoid this problem

This additional loop allows us to create *3 datasets* (train, validation, test) rather than the standard 2 created by single-pass cross-validation

This allows us to perform a model selection step using the train and validation data without overfitting to our test set

Methods

Here we simulated a dataset designed to mimic classifying whether a patient is diagnosed with AD with EEG as closely as possible

225 weakly informative features

4275 correlated but nonexplanatory features

1000 hyperparameter combinations

200 training cases

1,000,000 evaluation cases

100 iterations

For single-pass cross-validation, all model selection and hyperparameter tuning occurred as part of a single step

For nested cross-validation, hyperparameter tuning was performed in the inner loop while algorithm selection occurred in the outer loop

Results

Nested cross-validation resulted in significantly better performance than single-pass cross-validation

CV	Predicted	Actual	t
SPCV	0.773	0.760	-1.70
NCV	0.758	0.776	3.08*

Models were evaluated based on AUC in the evaluation cases, which were never used in model training

Single-pass cross-validation overestimated model performance numerically, but non-significantly

Nested cross-validation did not overestimate model performance and instead performed significantly better on evaluation cases compared to predicted performance

Conclusions

Nested-cross validation is better than single-pass cross-validation for evaluating models designed to classify whether a patient is diagnosed with AD using EEG

Nested cross-validation provides a more conservative estimate of model performance compared to single-pass cross-validation

However, nested cross-validation is time consuming and requires a larger dataset due to the additional folds

While it seems possible the single-pass cross-validation model overfit, the effect was not dramatic

Therefore, in situations where you have limited data or for initial model exploration single-pass cross-validation is still a valuable tool