PCA + Logistic regression

PCA across trials with one channel:

Data: 180 x 61 x 750 (sample size   x   channel   x   timepoints)

Preprocessing:

Regularized data using StandardScaler from Scikit-learn

Using only one channel (C3), so the data becomes (180 x 750)

Purpose of PCA:

Find the axis where these 180 points (each of 750 dimensions) have the most variance, project the data onto these axes. The transformed data are expected to be more discriminative.

Result: With PCA of 10 components, (180 x 750) is transformed into (180, 10). Feeding this into logistic regression gives Classification accuracy: 0.280556 / Chance level: 0.250000. (four classes)

Another linear classifier, LDA, gives slightly better result Classification accuracy: 0.313889 / Chance level: 0.250000

Alternative method using PCA:

Instead of PCA across different trials, I attempted to do PCA across channels and received a much better result.

Purpose of PCA:

By doing PCA across channels, I am looking for axes where channels have the most variance.

Result:
For each trial out of 180 trails, (61, 750) is transformed into (61, 10). Then it is flattened into 1 x 610 and feed into logistic regression, it gives Classification accuracy: 0.491667 / Chance level: 0.250000. This is 20% more accurate than the previous PCA method.

Common Spatial pattern:

Common spatial pattern is another form of PCA but focus on discriminating two classes.

$$w_c = \underset{w}{\max}\ w^\top \Sigma^{(c)} w \ \text{s.t.}\ w^\top \big(\Sigma^{(-1)} + \Sigma^{(+1)}\big)w = 1$$

In the above formula for CSP, the first half is the same as PCA, the second part gives a constraint that the sum of variance for both classes are constant. Therefore, maximizing one class means minimizing the other class.

Purpose of CSP:

Finding axes where variance of one class is maximized while the variance of another class is minimized.

Result:

CSP focuses on the variance of timepoints in a single trial. The covariance matrix of each class is the average of the trials belong to this class. CSP then finds the axis the maximize the variance among timepoints for one class and minimize that of the other class.

CSP with n_component=4 gives Classification accuracy: 0.800000 / Chance level: 0.250000 under 10 folds cross validation. This is 30% better than PCA across channels.