Grasp-and-Lift EEG Detection

Anirudhan J Rajagopalan, Michele Cerú

New York University

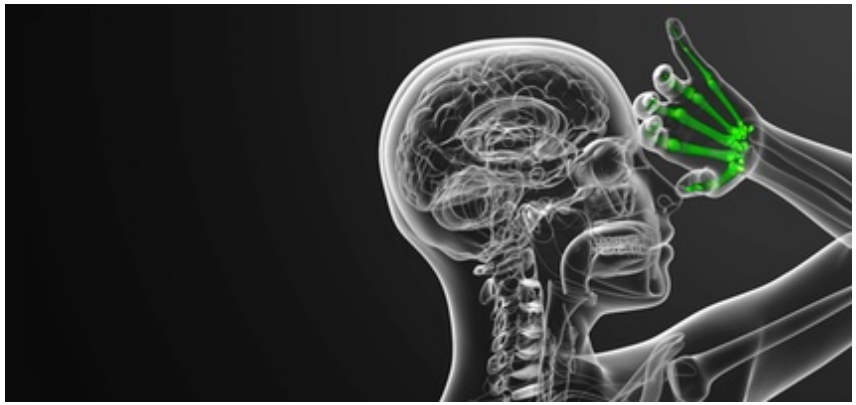
ajr619@nyu.edu; mc3784@nyu.edu

December 16, 2015

Project description

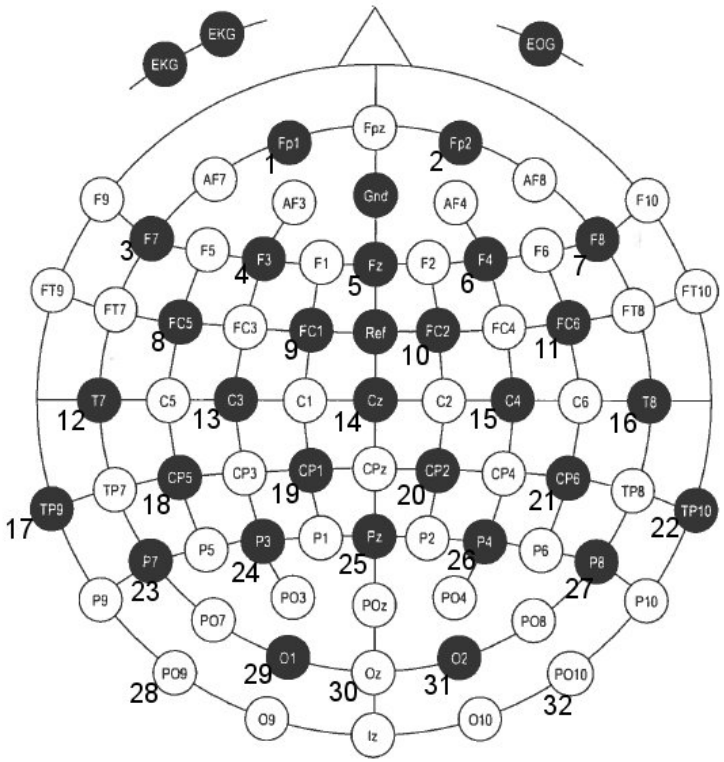
Identify hand motions from EEG recordings

- Goal: classify hand motions from its EEG signal data.
- Purpose: Brain-Computer Interface prosthetic devices for restoring a patient's ability to perform basic daily tasks.



Dataset

32 EEG signals:



- 30 Grasp And Lift series.
- Training data set: 96 files
- testing data set: 24 files
- Size: 1.5 Gb.
- 17985850 total number of samples
- $\sim 180k$ samples per subject
- sampling rate: $500Hz$
- Multi class classification: *Hand start*, *First digit touch*, *Start load phase*, *lift off*, *Replace*, *Both released*

Multilabel vs Multiclass

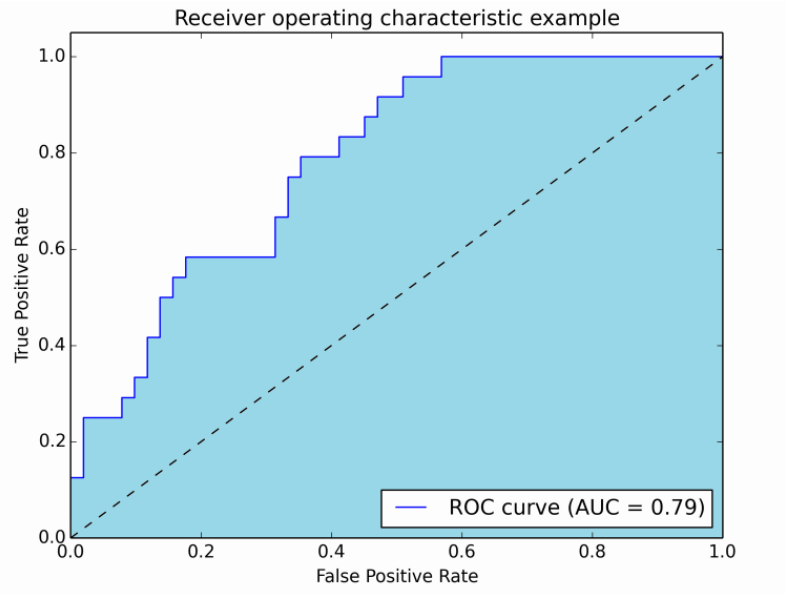
- Multilabel classification: each sample could have a set of different target labels
- Multiclass classification: each sample is assigned to one and only one label

```
subj9_series7_126708,0,0,0,0,1,0
subj9_series7_126709,0,0,0,0,1,0
subj9_series7_126710,0,0,0,0,1,1
subj9_series7_126711,0,0,0,0,1,1
subj9_series7_126712,0,0,0,0,1,1
subj9_series7_126713,0,0,0,0,1,1
subj9_series7_126714,0,0,0,0,1,1
subj9_series7_126715,0,0,0,0,1,1
subj9_series7_126716,0,0,0,0,1,1
subj9_series7_126717,0,0,0,0,1,1
subj9_series7_126718,0,0,0,0,1,1
subj9_series7_126719,0,0,0,0,1,1
```

Evaluation Criteria

Mean Column-wise Area Under the Curve (MCAUC): the mean of individual areas under the ROC curve for each predicted columns.

Sample ROC curve



Pipeline

Preprocessing

KDawn Filter: with hyper parameter 2-3-4

VLAD

Number of clusters: $2^3 \rightarrow 2^{15}$

PCA

number of components $= 0.9$

SVM - Linear & Gaussian

C and γ varies from 2^{-3} to 2^3

Performance optimization

- Preprocess: store the data for each component and use that in the next step of the pipeline
- VLAD: save intermediate states as bumpy binary files and use them for the other parts of the pipeline
- kmeans: inertia convergence criteria.

Pipeline Time

Total time $\sim 5h$

- Preprocessing: $< 1h$
- All other steps: $\sim 4h$

VLAD: $\sim 12s$ with 32 clusters and $1M$ local descriptors (with ubuntu dual core cpu, 3Gb ram)

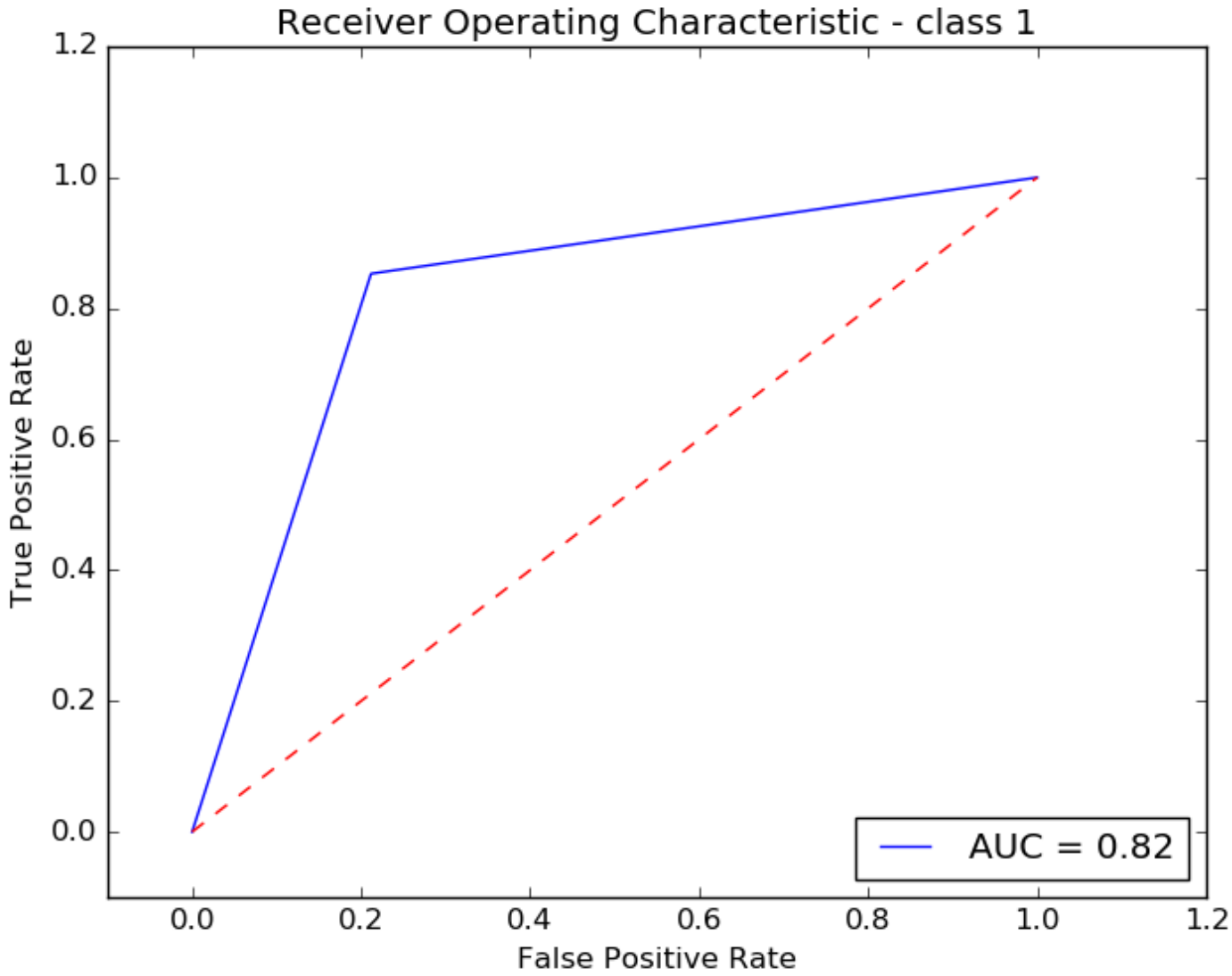
Table

γ	C	mean score
1	1	0.36215
2	1	0.3222
4	1	0.27447
0.25	2	0.41857

Table: N components=3, $k = 1024$

AUC

Number of clusters: $k = 1024$:



Obstacles

- Domain knowledge.
- Preprocessing steps
- Implementing vlad that is compatible with scikit learn's pipeline.
- Brutal HPC clusters.

Future work

- Preprocess using filterbank which helps pick up low frequency but important features.
- Training multiple models for the same problem.
- No future data rule.

References

- https://www.kaggle.com/c/grasp-and-lift-eeg-detection/forums
- https://github.com/alexandrebarachant/Grasp-and-lift-EEG-challenge
- http://www.gipsa-lab.grenoble-inp.fr/~bertrand.rivet/references/Rivet2009a.pdf
- https://lear.inrialpes.fr/pubs/2010/JDSP10/jegou_compactimagerepresentation.pdf
- http://www.eecs.tufts.edu/~dsculley/papers/fastkmeans.pdf

Acknowledgements

- Zaid Harchaoui
- Alexandre Barachant and Rafal Cycoń

The End