Epileptic seizure detection

Alex YE

https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition

EEG signals: **1 value** over **23.5 seconds (4097 points)** for each of the **500 patients**

Reduced to 1 second signals: **178 points** for **11500 entries**

The response variable is y in column 179, the Explanatory variables X1, X2, ..., X178

y contains the category of the 178-dimensional input vector. Specifically y in {1, 2, 3, 4, 5}:

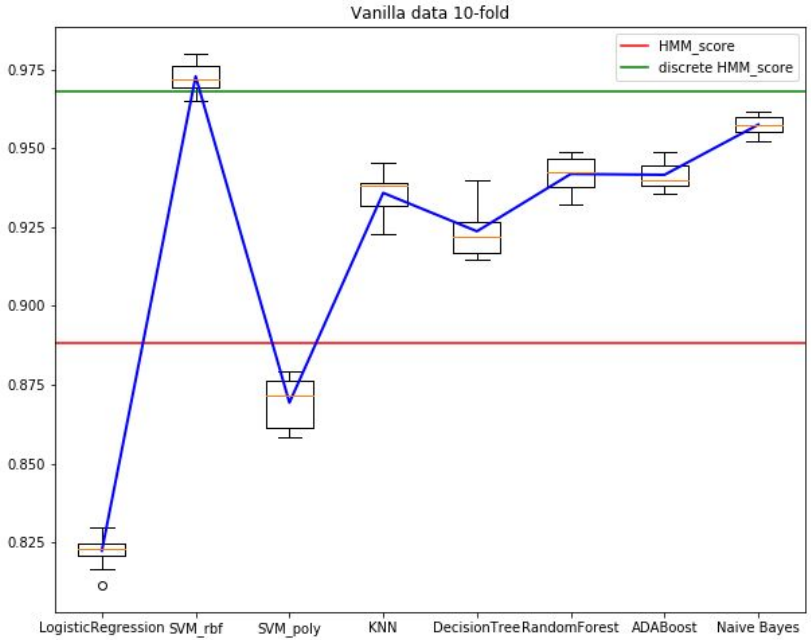
5 - eyes open, means when they were recording the EEG signal of the brain the patient had their eyes open

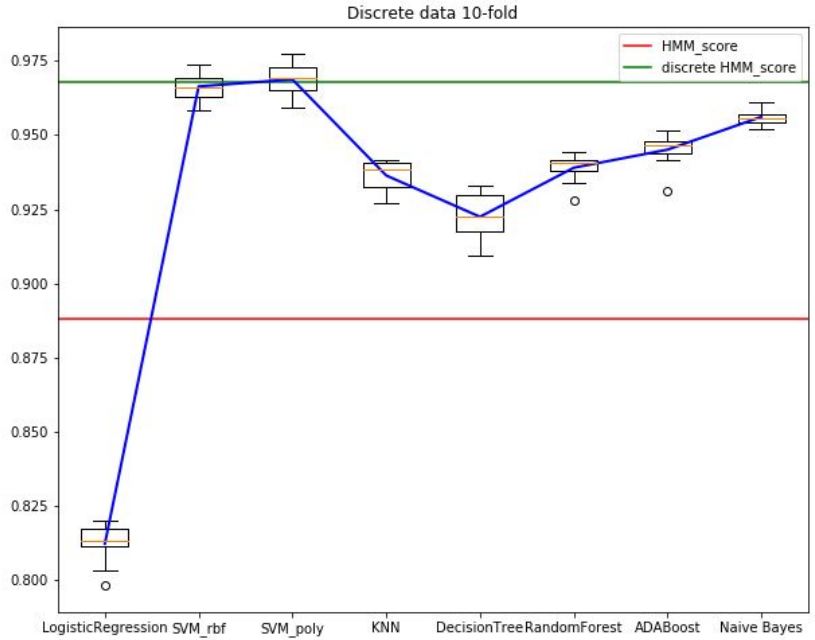
4 - eyes closed, means when they were recording the EEG signal the patient had their eyes closed

3 - Yes they identify where the region of the tumor was in the brain and recording the EEG activity from the healthy brain area

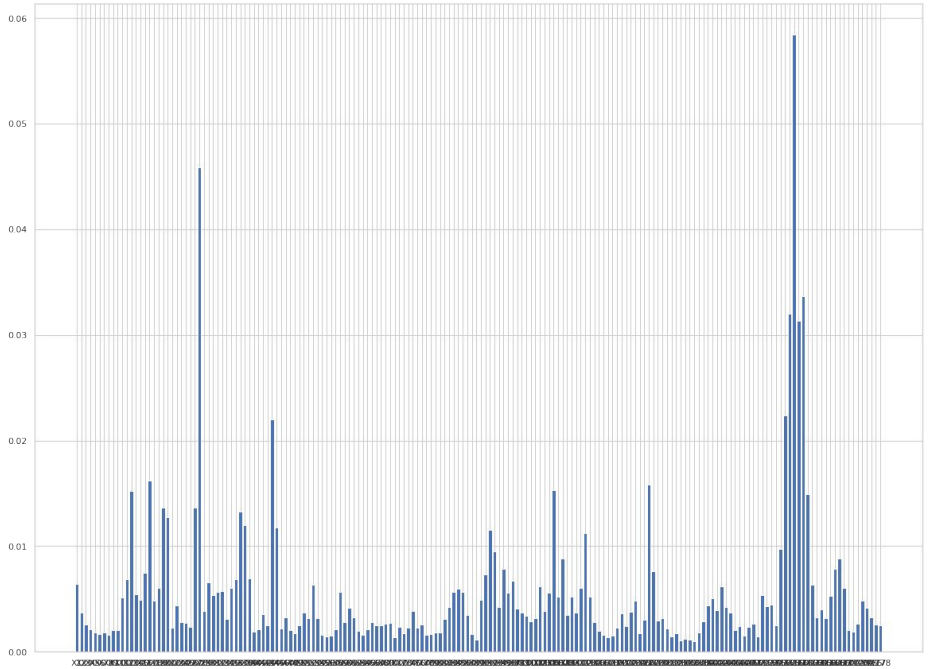
2 - They recorder the EEG from the area where the tumor was located

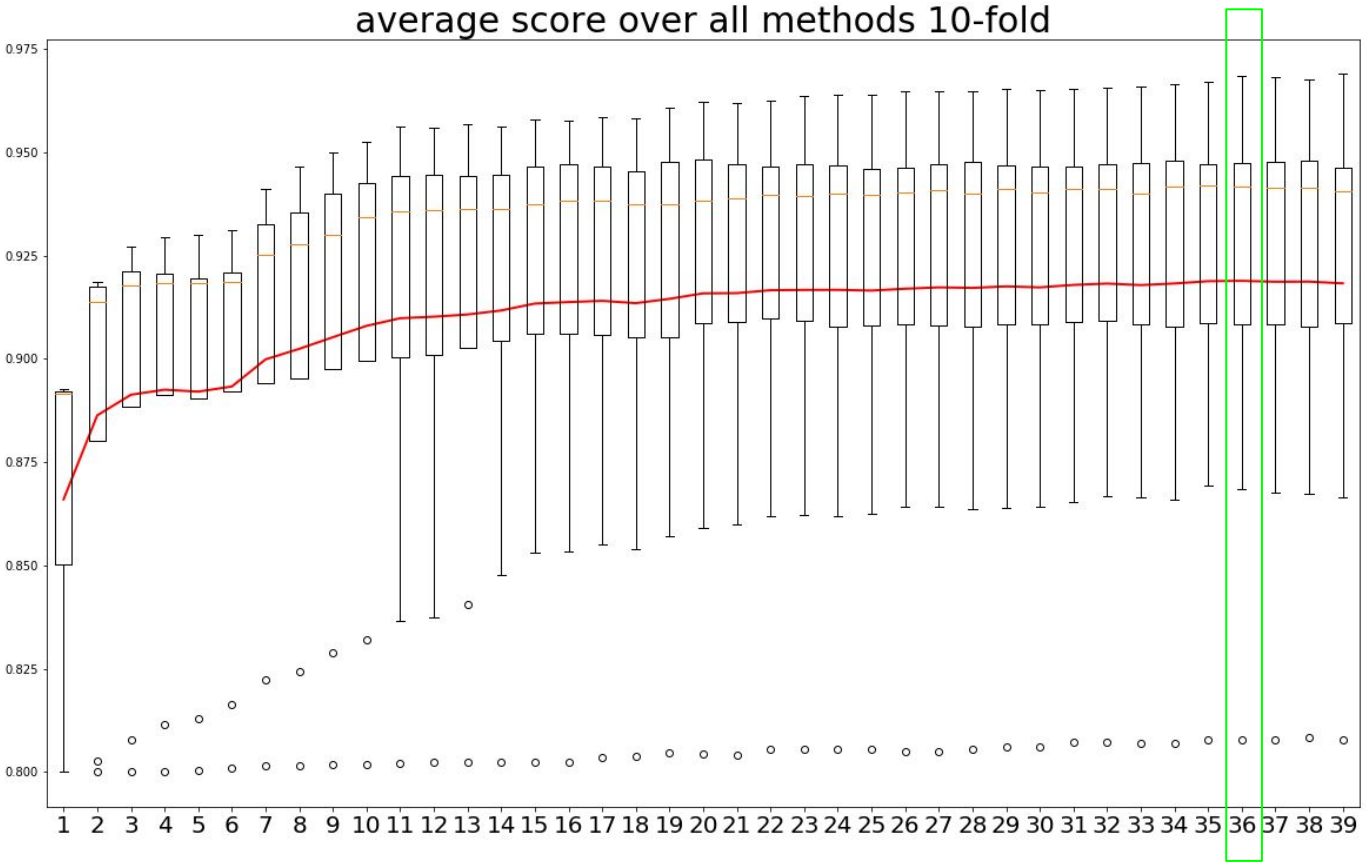
1 - Recording of seizure activity

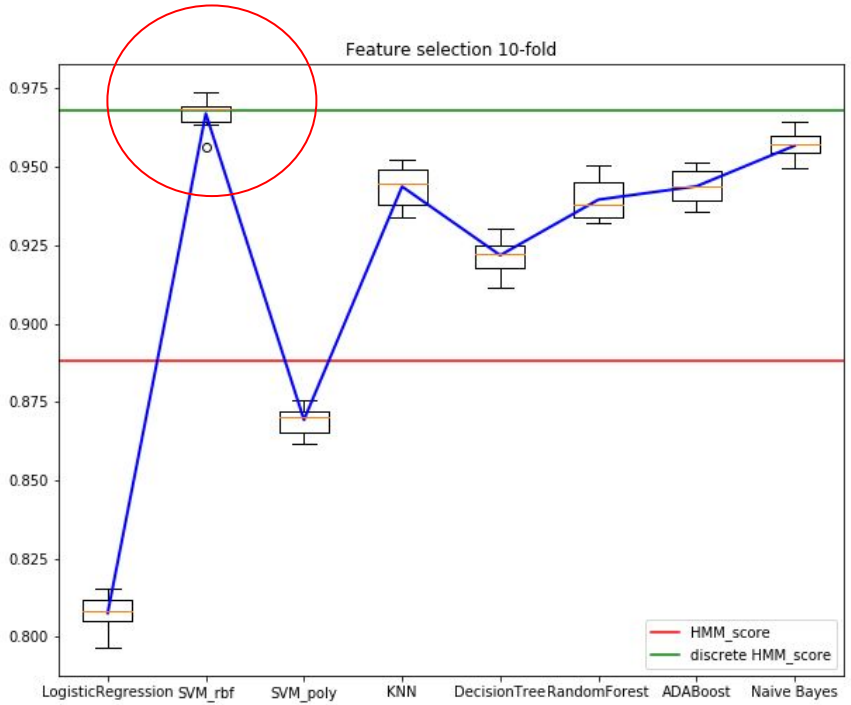

Vanilla data 10-fold
HMM_score
discrete HMM_score
0.975
0.950
0.925
0.900
0.875
0.850
0.825
LogisticRegression
SVM_rbf
SVM_poly
KNN
DecisionTree
RandomForest
ADABoost
Naive Bayes


Discrete data 10-fold
HMM_score
discrete HMM_score
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
LogisticRegression
SVM_rbf
SVM_poly
KNN
DecisionTree
RandomForest
ADABoost
Naive Bayes

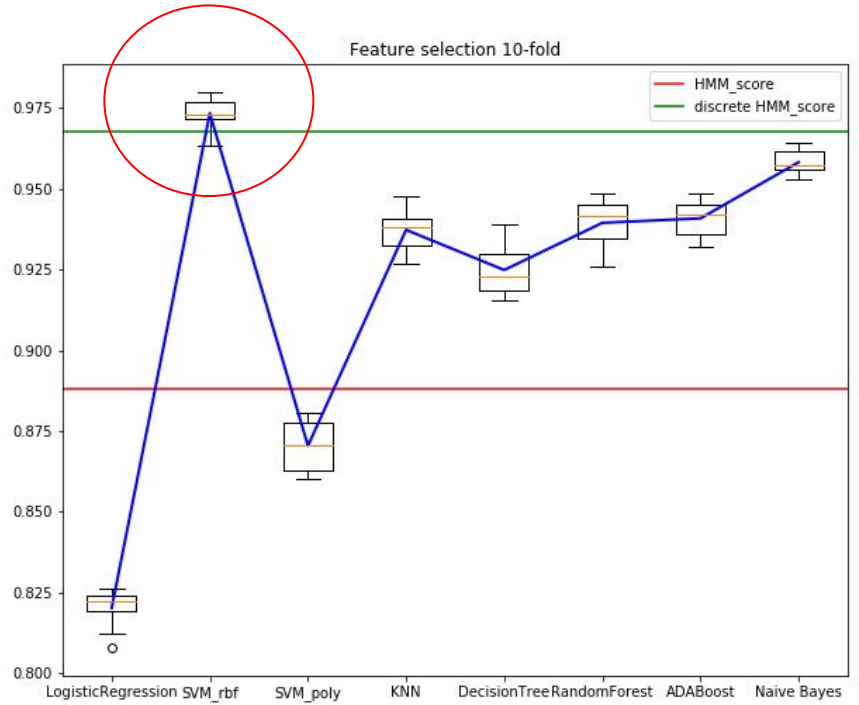
Feature importance (random forest)




average score over all methods 10-fold
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
nb feature



Best features

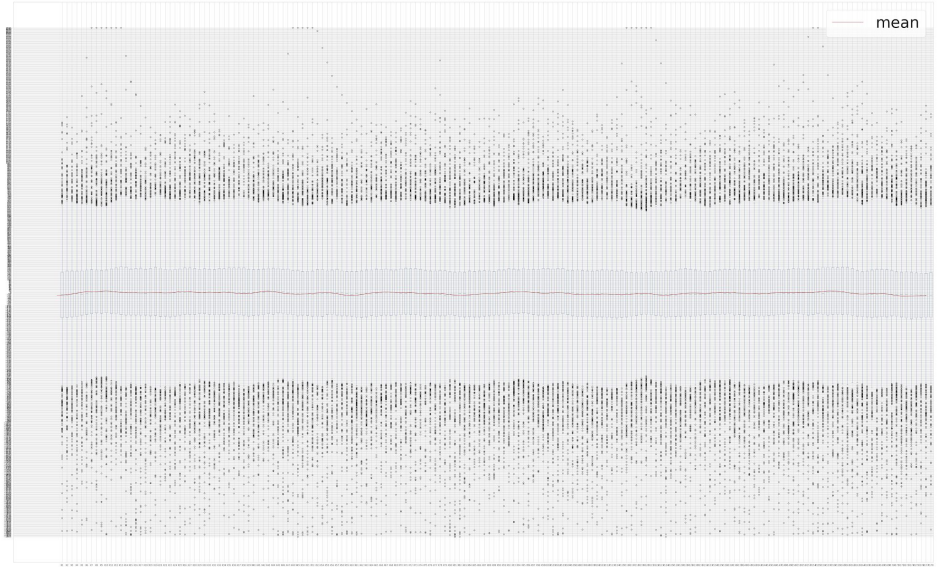


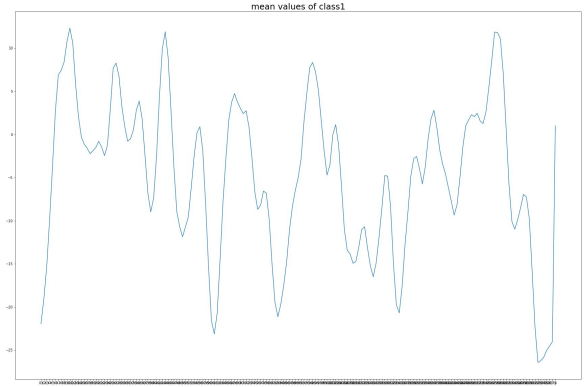
Worst features

one more thing

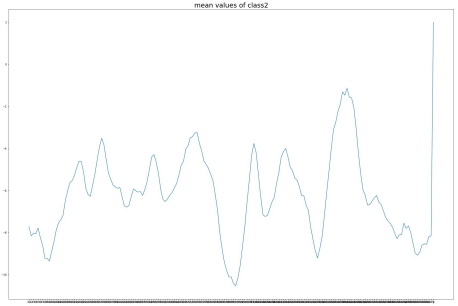
Features distribution

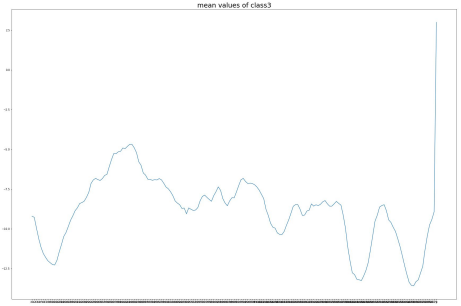
We divided and shuffled every 4097 data points into 23 chunks, each chunk contains 178 data points for 1 second, and each data point is the value of the EEG recording at a different point in time. So now we have 23 x 500 = 11500 pieces of information(row), each information contains 178 data points for 1 second(column).

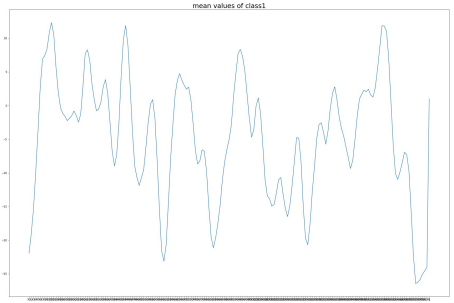


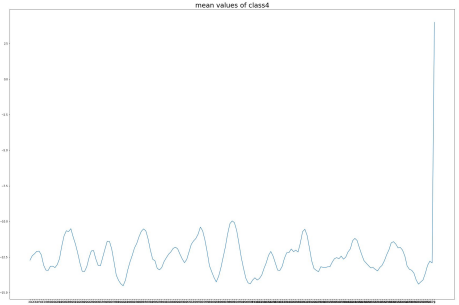


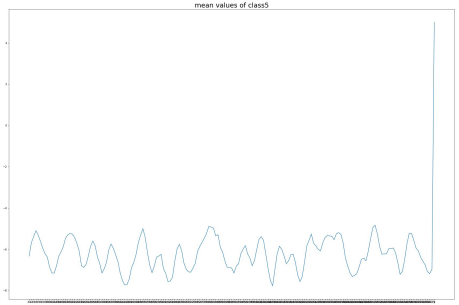
Mean curves


mean values of class2


mean values of class3


mean values of class1


mean values of class4


mean values of class5