Seminar Hands-on BCI implementation

Session 5: Feature preconditioning and extraction (t-values and relief algorithms)

Load data file to workspace (results from session 3 and 4)
load ecogStruct3.mat
load zScoredData.mat
load epoch2.mat is needed to find out which trial belongs to which class (extension or flexion), stored in epoch.label

t-Values, to choose the best features (using the z-scored data)

Subsets

Features for classification are often selected from the training data set. In avoidance of being to specific features are typically determined from a subset of the whole training data.

Create subsets (ratio of entire data, i.e. 0.9 means 90 %)
[subSet1, subSet2] = createSubsets(dat,epoch,0.9);
subSet1 will contain 90% of all finger flexion trials
subSet2 will contain 90% of all finger extension trials

TASK 1 (1 pt):
Have a short look at the createSubsets function to fully understand how the subsets are created. This function can only be used for the finger flexion/extension classes. Why?

TASK 2 (2 pt):
T-Values:
We will use t-values to compare the subsets.
Calculate variance and mean for each feature

Calculate sample size for each subset

Then calculate the t-values using the following formula:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{(N_1-1)S_1^2 + (N_2-1)S_2^2}{N_1+N_2-2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

```
Plot feature space to manually select features
plotFeatures(abs(tVals),ecog.selectedChannels,nFreq)
```

TASK 3 (2 pt):
First have a look at the plotFeatures function to understand how this plot was
created.
What information can you get from this plot? What do the colors mean?
Why do we plot the absolute tVals?
Make sure you write your observations down, as we will need them next week for
the classification.

you can compare your results to the anatomical image:

```
Anatomy = imread('GP33_anatomy_100electrodes.png');
figure
imshow( Anatomy );
```

Relief algorithm, an alternative way to choose the best features (also using the z-scored data)

Relief algorithm is a supervised feature selection method based on
k-nearest neighbor method

```
% Create subsets (ratio of entire data, i.e. 0.9 means 90 %)
[subSet1, subSet2] = createSubsets(dat,epoch,0.9);
subSet1 will contain 90% of all finger flexion trials
subSet2 will contain 90% of all finger extension trials

Create one feature vector
reliefData = [subSet1; subSet2];

Create label vector
reliefLabel = [repmat('FL',size(subSet1,1),1);
repmat('EX',size(subSet2,1),1)]; % FL = flexion;  EX = extension

Number of compared neighbors
k = 20;

Apply relief
[rank, weight] = relieff(reliefData,reliefLabel,k);

Weights are assigned from -1 (unimportant) to 1 (important). For a better
visualization this range is rescaled.
weight = weight + abs(min(weight));

Plot feature space to manually select features
plotFeatures(weight,ecog.selectedChannels,nFreq)
```

TASK 4 (1 pt):

What does the relieff function do? What information is stored in rank and
weight?
What information can you get from this plot? What do the colors mean?
Are these results comparable to the results from the t-values?