Report Practical Exercise 02

Task 1:

1.a)

The Dice coefficient measures how big the overlap between the predicted segmentation and the ground truth is. The formula is “2 \* intersection of A and B / A + B”. This can turn into “2\*TP / TP+FP + FN+TP” for Boolean data like in our case, which equivalent to the F1 score.

The recall compares the amount of correctly labelled area to the area that should be labeled as positive (TP /TP+FN).

The precision compares the amount of correctly positively segmented area to the total positively segmented area (TP/TP+FP).

The F1-score and therefore also the Dice score is the harmonic mean of precision and recall.

TODO: Pros/ Cons

1.b) Code explanation?!

1.c) Code explanation?!

1.d) Bonus optional

Task 2:

2.a)

We chose KMeans ... TODO

2.b)

We chose Gaussian Matrix Modelling because ... TODO

Task 3:

TODO

Task 4:

TODO: ergibt das mit dem intensity plot so Sinn?

4.a The most intuitive, non-learning-based approach would be to take thresholds and segment based on those and the Intensity Density Plot. Thresholds could simply be put around the peaks (local maxima) of the graph with the borders between thresholds being in the local minima. Rough estimates of the thresholds for the graph in 2 could then be [-100; 90], (90, 225], (225, 400], (400, 550] for background, CSF, white matter and grey matter respectively. These local minima would be calculated by the code automatically.

4.b)

Qualitative: (looking at images)

Especially for the worst prediction on the test set, even an untrained eye can see, that parts of the brain that should be labelled remain unlabelled as background. Even in the best performing prediction on the test set there are black zones, that should be labelled differently especially in the axial and sagittal view. Nonetheless the overall image, especially for the best prediction largely resembles the ground truth.

Quantitative: (scores & numbers)

While the maximum dice score of 0.74 would be good if it were an average of the whole validation set, the lowest achieved dice score of 0.46 is too bad to be useful in a real-world scenario.

It might still be useful as a basis for further human segmentation.