SPACE SCIENCE: ASSIGNMENT 1 2024

This assignment is broken up into three parts, A, B and C. In this assignment you will learn more about the fundamentals of space science and signal processing. You will also learn how to implement/use some Machine Learning algorithms that are frequently used in space science.

The Data:

- The data that you will be using for this assignment will be the Gauteng dataset that is provided. Note for Belief propagation you will not use this data.
- When using the data for part B break the data up into two equal parts, randomly
    - 50% of the data should be for training
    - 50% of the data should be for testing
    - Each group has to use a unique seed for the random splitting of the data.

Part A (30%):

Implement the Belief Propagation Algorithm, the final resulting graphs need to be Smooth. See Assignment 1 - Part A and B.ipynb for the details.

Part B (50%):

- Read the provided papers:
    - **[1]** T. L. Grobler, W. Kleynhans and B. P. Salmon, "Empirically Comparing Two Dimensionality Reduction Techniques – PCA and FFT: A Settlement Detection Case Study in the Gauteng Province of South Africa," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium , Yokohama, Japan, 2019, pp. 3329-3332.
    - **[2]** T. L. Grobler, W. Kleynhans and B. P. Salmon, "Sequential Classification of MODIS time-series," IGARSS 2012 - 2012 IEEE International Geoscience and Remote Sensing Symposium , Munich, Germany, 2012, pp. 6236-6239.
- Replicate the FFT comparison results of paper [1] using the provided Gauteng dataset.
- All the steps that are required to replicate the FFT comparison results are provided in the paper.
- In particular you should replicate the graphs in Figure 2 of paper [1] and report the results in your own words.
- In addition, implement a Logistic Regression and a Naive Bayes classifier using the FFT features for each band in Figure 2. Report the classification accuracies of each classifier and band in a bar graph. Use 50% of the data for training and 50% for testing. Use the scikit-learn function *train_test_split* with a random seed of 42 to split your data. For the hyperparameter C of logistic regression use 1e5.

- You may use scikit-learn's implementations for logistic regression and naive bayes and numpy's FFT implementation.
- Note: You **DO NOT** have to replicate the PCA results. The results reported in Fig. 3 are incorrect, the PCA and FFT results should be swapped around.
- Now implement a per band time-varying sequential model as described in paper [2] (from first principles). To build the pdfs you need for the model you may use scikit-learn's GMM implementation (PS use only one GMM component to estimate a pdf). Use only 50% of the data to construct the model (training data) and 50% to test the model's performance. Use the same split as before. You may disregard any observations that were made after the first 8 years of observations when you train your models.
- Now use the SPRT test (not eq 2 in paper [2] but the approach discussed in class - use LLRs see caption of Figure 4 in paper [2]) and the time-varying model you created to discern between vegetation and settlement pixels (apply it to the test data set). Add the per band accuracies so obtained to the aforementioned bar graph.
- Note your results will not match the results of paper [2] exactly, as the time-varying model in the paper was built using all the data and used KDE to construct the pdfs required.
- Repeat the experiment for paper 1 and 2 ten to 20 times and report the standard deviations so obtained in whatever way you see fit (please use a different random seed for each experiment). Report the results of both the single and multiple run experiments.
- Interpret the results and provide a final conclusion.

<u>PART C (20%):</u>

For this part of the assignment you need to implement the Radix-2 DIT FFT algorithm. Please complete the notebook: fft_implementation_assignment.ipynb. Note your notebook will be graded. Note, you NEED NOT simplify the butterfly calculations by re-using the twiddle factors: https://cnx.org/contents/qAa9OhlP@2.44:zmcmahhR@7/Decimation-in-time-DIT-Radix-2-FFT

<u>What needs to be submitted?</u>

- All the code used to produce the results from part A, B and C.
- A report that reports on the results obtained from part A, B and C explaining the results, algorithms and showcases that you understand the topics.
- Plagiarism form (digital signature from each student)

<u>What needs to be in the Report?</u>

- Reporting on Part A, B & C.
- Belief Propagation explained & discuss the obtained results,
- Remote Sensing with Machine Learning explained & discuss the obtained results,
- Fourier Transform, DFT and FFT explained & discuss the obtained results,
- The content of the report includes but is not limited to,
    - Introduction
    - Definitions,

    - Formulas,
    - Pseudocode,
    - Figures
    - Conclusion
- Only the report will be graded (the code will in some instances be graded but will count much less than the report), but proper working code has to be provided as it will be compiled and compared to the results presented in the report.
- The report must have a unified look and feel.
- No collusion between groups, it has to be each group's own work.

Report Format and particulars?

- Format: IEEE (two-column) or ACM,
- PDF-version,
- Proper IEEE/ACM referencing.
- Number of pages (including references): 8

TIPS:

- Less is more, if you write a report and you can say blah , please don't try and say blah blah blah.
- Do not over explain as this will reduce the mark for your report.
- Use your own words and make use of reliable resources.
- It is recommended that you divide part A, B and C up between the group members (to reduce the workload) and work together on writing the report.
- Please note that during a theory test you may be examined on any part of this assignment so please make sure you understand the entire assignment.
- If a group member does not contribute 100% to the assignment, this can be reflected in peer-review (at the end of the third term), and this will reduce their mark according to the peer-review criteria.