CS675: Computer Vision Assignment 2

Due date: Apr 3rd at 11:59pm.
Topic: Classification
Assignment Title: Classification using traditional machine learning methods

Instructions:

In this assignment we will test different classifiers over a few datasets.

Classifiers: The classifiers you should use for all items below are:

1. Nearest Neighbors
2. Linear Discriminant Analysis
3. Logistic Regression
4. SVM
5. MLP

Many open-source implementations for these algorithms exist. Scikit-learn (or sklearn) is a good place to start where many of those algorithms are implemented and it provides a nice tutorial on how to use them (see, https://scikit-learn.org/stable/).

Datasets: In this activity you use both synthetic and real datasets.

Task 1: In this problem we will generate a synthetic dataset, train and test the 4 classifiers discussed above, compare their performances and plot their decision boundaries.

1. Create a 2-class, 2-dimensional dataset where the inputs x are drawn from Gaussian distributions. Being a two-class dataset, the data points x from each class should be drawn from its Gaussian. Labels should also be attributed to each sample accordingly. Following this procedure, generate training and test sets.
2. Train the 4 classifiers discussed above. These classifiers may have hyperparameters that may require tuning. For instance, the Nearest Neighbors have a k (number of neighbors) that needs to be selected. A good practice is to use cross-validation techniques to tune such parameters. Perform stratified cross-validation to tune the hyperparameters (https://scikit-learn.org/dev/modules/cross_validation.html).
3. Compare the performance of the different classifiers in terms of accuracy.
4. Display their confusion matrices.
5. Plot the decision boundary for each classifier (https://scikit-learn.org/dev/modules/generated/sklearn.inspection.DecisionBoundaryDisplay.html).

Task 2: In Task 2 we will do the same approach but now with a real image dataset. For this we will use the scikit-learn digits dataset: https://scikit-learn.org/stable/auto_examples/datasets/plot_digits_last_image.html . Perform the same steps of Task1 except for the step 5 (decision boundary). This cannot be as easily done here since the dimension of our data points is 64 (pixels).

Task 3: Perform the same as in Task 2 but using CIFAR-10 dataset. (check: https://www.tensorflow.org/datasets/catalog/cifar10)

Submission Instructions:

- **Format:** Submit your work as a **Jupyter Notebook (.ipynb)** file, ensuring all code cells are executable. Additionally, export the executed notebook as a **PDF or HTML report**, where all outputs and results are visible.
- **Structure of Submission:** Each task (**1, 2 & 3**) in your assignment should follow this structured approach:

    a. Introduction and objectives (briefly describe the purpose of the assignment)
    b. Implementation with justification (explain your solution and justify any design decision)
    c. Results and analysis (provide error metrics and confusion matrices for the classifiers)
    d. Discussion of findings (focusing on the performance of each classifier)

    At the end of the report, provide a conclusion summarizing key findings and reflecting on the effectiveness of the approaches used, highlighting their strengths and limitations.

    e. Conclusions

Evaluation Criteria:

Your assignment will be evaluated based on the following criteria:

1. **Introduction and objectives (5%)**
    - Clearly state the purpose of the assignment.
    - Explain what you are doing and why it is important. Provide context or background information if necessary.
2. **Implementation with justification (25%)**
    - Present a well-structured and efficient implementation. Ensure proper documentation and readability of the code.
    - Justify every critical decision made in the code. Explain the logic behind your approach, choice of algorithms, tools, or frameworks.
3. **Analysis of the results (30%)**
    - Explain the results in relation to the objectives. Present them clearly, using plots, graphs, or explanations as needed.
    - Highlight any patterns, trends, or key insights from the findings.
4. **Discussion of findings (40%)**
    - Critically evaluate the results and their implications.
    - Compare findings with expected outcomes (if applicable).

- Discuss any limitations or challenges faced during the process. Provide suggestions for improvements or future work.

Good luck and have fun!