ML Assignment 4

Submission Deadline: 23:59 , Sunday ,October 22 , 2023
No late submissions will be accepted.

Objective:

The objective of this assignment is to provide students with hands-on experience in performing classification tasks using dimensionality reduction techniques: Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) on the CIFAR-10 dataset. Students will gain insights into the effectiveness and differences in using PCA and LDA for classification.

Dataset:

CIFAR-10 is a widely-used dataset in the field of computer vision. It contains 60,000 32x32 color images across 10 different classes, with 6,000 images per class. The dataset is divided into a training set of 50,000 images and a test set of 10,000 images. Each class represents a specific category of objects, providing a diverse range of images for training and evaluation.

The 10 classes in CIFAR-10 are as follows:

- Airplane
- Automobile
- Bird
- Cat
- Deer
- Dog
- Frog
- Horse
- Ship
- Truck

Data Preparation:

a. Download the CIFAR-10 dataset.
b. Preprocess the dataset
c. Randomly select 80% train set and 20%test set. Train on the training set and test on the test set.
d. Prepare data matrices and structures to hold data and labels

Task 1: Classification Using Principal Component Analysis (PCA):

a. Implement PCA to compute PCA Basis of each Class.
b. Choose an appropriate number of principal components to retain for each class.
c. Write a projection function to project data on PCA basis
d. Write a backprojection function to bring back data to original space and visualize data.
e. Write a function for computing error between projected and back-projected data
f. Write a function to compute labels of test images using minimum error criterion.
g. Evaluate the models' performance using accuracy, precision, recall, and F1-score on the testing set.

Task 2: Classification Using Linear Discriminant Analysis (LDA):

a. Implement LDA to compute the LDA basis for the CIFAR-10 dataset for 10 classes.
b. Write a function to project data to the LDA space. Visualize projected data in three dimensional space.
c. Train a small neural network for classification on the reduced dataset using LDA.
d. Evaluate the models' performance using accuracy, precision, recall, and F1-score on the testing set.

Task 3: Classification using Autoencoders

a. Train an autoencoder for each class in CIFAR 10. Visualize output of the autoencoder as an image. Choose appropriate size of neurons in input and output layers with a single hidden layer.
b. Write a function to measure the reconstruction error between input and output or autoencoder.
c. Write a function to find the label of a test image using autoencoder based on minimum error criterion.
d. Evaluate the models' performance using accuracy, precision, recall, and F1-score on the testing set.

Task 4: Comparative Analysis:

a. Compare the performance of the PCA-based model, LDA-based model and autoencoder based model in terms of accuracy, computational efficiency, and other relevant metrics.
b. Discuss the trade-offs and advantages of using PCA, LDA and AutoAE for dimensionality reduction in classification tasks.

Task 5: Bonus Task

Train Autoencoder using label supervision. You have to think of a strategy to incorporate label information into the autoencoder training stages.

Conclusion:

Write a conclusion summarizing the findings and insights gained from this assignment, including the comparison of PCA and LDA in the context of the CIFAR-10 dataset.

Submission Guidelines:

- The code should be well-documented and organized with appropriate variable names and
- comments.
- Ensure that the plots are clear and labeled appropriately.
- Submit well-commented code of Python .
- Write the Report explaining the results.
- Students should submit a report detailing their implementation, results, and analysis.
- The report should include visualizations (e.g., graphs, charts) where necessary to support their analysis.
- Students are encouraged to experiment with different hyperparameters and techniques to enhance their understanding of PCA and LDA.
- Failure to follow the correct naming convention will result in the assignment not being marked.

Assignment Naming Convention:
● The assignment must be named in the following format: firstname_rollnumber.zip. Example: Fatima_MSDS18089.zip

Assignment Submission Contents:

The assignment submission in the zip file should contain the following items:

- **Python Script (.py):** Name: firstname_rollnumber.py
- **Jupyter Notebook (.ipynb)**: Name: firstname_rollnumber.ipynb
- **Assignment Report (.pdf):** Name: firstname_rollnumber_report.pdf