ECE 520.438 & 520.638: Deep Learning
Homework 1
Spring 2024

Grading: You will be graded based on the code you develop, plus your homework report summarizing your findings. The code you develop can be in any language you prefer and you are allowed to use library functions. If possible, please write your report using LaTeX.

1 Part 1 (20 points)

Answer the following questions.

1. What do we mean by hand-crafted features? Give at least three examples of hand-crafted features.
2. What do we mean by learned features? Give at least three examples of leanred features.
3. Briefly discss some of the advantages and disadvantages of hand-crafted features compared to learned features.
4. What are two ways in which one can formulate PCA?
5. The LDA reduces dimensionality from the original number of features to how many features?
6. What some of the limitations of LDA?
7. Give two drawbacks of PCA.
8. Many features in computer vision are represented in terms of histograms. Given two histograms, what are some distance metrics that we can use to compare them? Give at least three examples.
9. Why does ℓ_0-norm capture sparsity?
10. Why do we use ℓ_1-norm to approximate ℓ_0-norm?
11. What are some disadvantages of k-means clustering?
12. What is the difference between Nearest Neighbor algorithm and k-Nearest Neighbor algoritm?
13. Breifly describe how visual bag of words features are extracted.
14. Briefly describe cross-validation.
15. What is the difference between sparse coding and dictionary learning?

2 Part 2: Face Recognition: k-NN (40 points)

In this excercise, you will implement and evaluate the k-Nearest Neighbor algorithm that we studied in class.

2.1 Extended YaleB dataset

1. The original images in the Extended YaleB dataset have been cropped and resized to 32×32. This dataset has 38 individuals and around 64 near frontal images under

different illuminations per individual. Sample images from this dataset are shown in Figure **??**. Download the file **YaleB-32x32.mat** from the course locker. This file contains variables **'fea'** and **'gnd'**. Each row of **'fea'** is a face and **'gnd'** is the label. Randomly select ($m = 10, 20, 30, 40, 50$) images per individual with labels to form the training set, and use the remaining images in the dataset as the test set. Apply the k-NN algorithm (with $k = 1$) on each of these five splits and record the corresponding classification errors. Use the Euclidean distance metric, i.e., $d(x, y) = \|x - y\|_2$. The classification error rate, E is defined as follows

$$E = \frac{\sum_{i=1}^{n} \mathbb{1}[\hat{l}(x_i) \neq l(x_i)]}{n} \times 100,$$

where n is the number of test samples, $\hat{l}(x_i)$ is the classification of the ith observation from the test, and $l(x_i)$ is the true class of that observation. Plot the *Classification Error Rate vs Number of Trainings Samples* curves on a single figure. Summarize your findings.

Figure 1: Samples images from the Extended YaleB dataset.

2. Repeat the above procedure for $k = 2, 3, 5, 10$ and plot the error rate E against k. Does the error rate decreases with k? Should the error rate always decrease with k? Plot the k nearest neighbors of some of misclassified samples.

3. Let $k = 3$ and select $m = 30$ images per individual with labels to form the training set and use the remaining images in the dataset as the test set. Replace the distance metric with $\|x - y\|_p, \quad p = 1, 3, 5, 10$ and plot the error rate against p. Does the distance metric effect the error rate?

4. Instead of using the pixel intensities as features, extract the LBP and HOG features from the images. Repeat step 3 with $p = 1, 2$. What are the error rates corresponding to pixel intensities, LBP and HOG features?

5. What is the lowest error rate you achieved in this exercise?

2.2 Validation set

Randomly sample 20 images per individual to form a test set. Use the remaining data to form the training set. Appropriately further divide the training set into a new training set and a validation set. Use the validation set to optimize the parameters (i.e k and p) of the k-NN algorithm. Please use the pixel intensities for conducting experiments in this exercise. What is the lowest error rate you achieved in this exercise? What are the corresponding values for k and p?

3 Part 3 - Face Recognition: Other algorithms (40 points)

In this part, you will implement and evaluate the four basic face recognition and image classification algorithms that we studied in class. These algorithms are Eigenfaces (PCA), Fisherfaces (LDA), Support Vector Machine (SVM), and Sparse Representation-based Classification (SRC) on the Extended YaleB dataset. You will need to follow the same procedure as in 2.1.1 and summarize the findings across these four algorithms i.e randomly select ($m = 10, 20, 30, 40, 50$) images per individual with labels (from YaleB) to form the training set, and use the remaining images in the dataset as the test set. Apply the each of the four algorithm on each of these five splits and record the corresponding classification errors. Use the Euclidean distance metric, i.e., $d(x, y) = \|x - y\|_2$. The classification error rate, E is defined as follows

$$E = \frac{\sum_{i=1}^{n} \mathbb{1}[\hat{l}(x_i) \neq l(x_i)]}{n} \times 100,$$

where n is the number of test samples, $\hat{l}(x_i)$ is the classification of the ith observation from the test, and $l(x_i)$ is the true class of that observation. Plot the *Classification Error Rate vs Number of Trainings Samples* curves for each algorithm. Summarize your findings.

Note that, for LDA, there are at most $c - 1$ nonzero generalized eigenvalues and, so, an upper bound on the dimension of the reduced space is $c - 1$, where c is the number of individuals.