ITU Computer and Informatics Faculty
BLG 454E Learning From Data, Spring 2018
Homework #3
Due May, 3 2018 10pm

1. (50 pts.) [**PCA**]

   (a) (3 pts.) What are the main motivations for reducing a dataset's dimensionality?

   (b) (3 pts.) How can you evaluate the performance of a dimensionality reduction algorithm on your dataset?

   (c) (2 pts.) What do you say about the performance of PCA in Figure 1 in terms of classification?

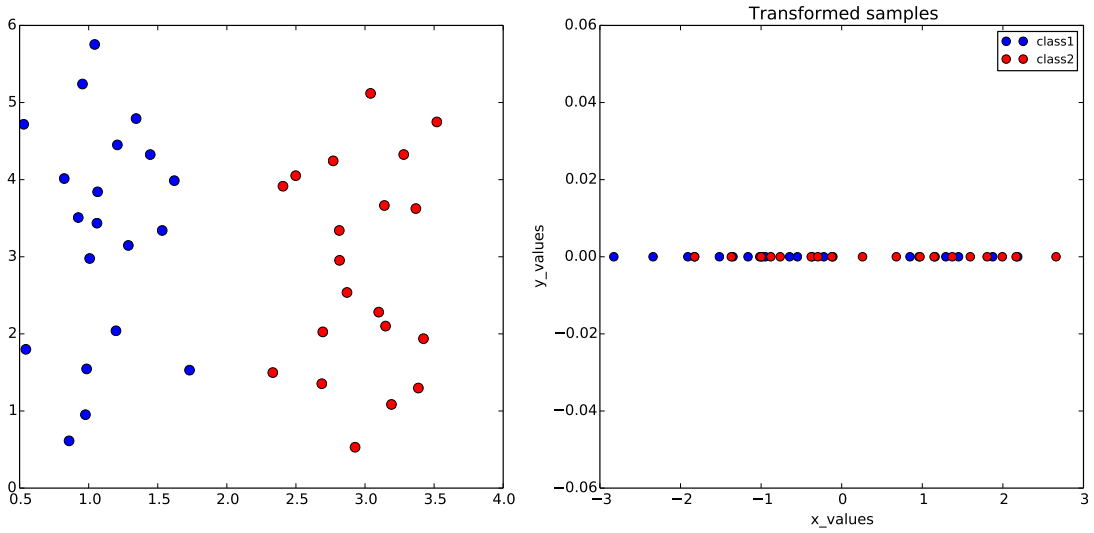

Figure 1: Original data in 2D (left) and its transformation in 1D (right).

   (e) (40 pts.) Implement a PCA projection on given the data.txt. The last attribute of the data.txt is the class label, range from 0..9.

   - Use the covariance matrix Σ to calculate the PCA components
   - Plot the transformed data points in 2D as shown in Figure 2. You need to use annotate() like function to write text(class label) at each randomly selected 200 data points
   - Give all your plots in your report.

   You are allowed use built-in functions to compute **eigenvalues** and **covariance** matrices. However, you are not allowed to use built-in function directly implements PCA.

2. (50pts.) [**SVD**] You are going to look at compressing the given RGB image, data.jpg, through computing the singular value decomposition (SVD). Each channel (red, green, blue) has 1537×2500 pixels which is a 1537×2500 matrix A.

   - (35 pts.) Find the SVD of A (one for each channel).
   - (15 pts.) Display the original image and image obtained from a rank (term) of 1, 5, 20, 50 SVD approximation of A as shown in Figure 3.
   - Give all your plots in your report.

   You are not allowed to use built-in function directly implements SVD.

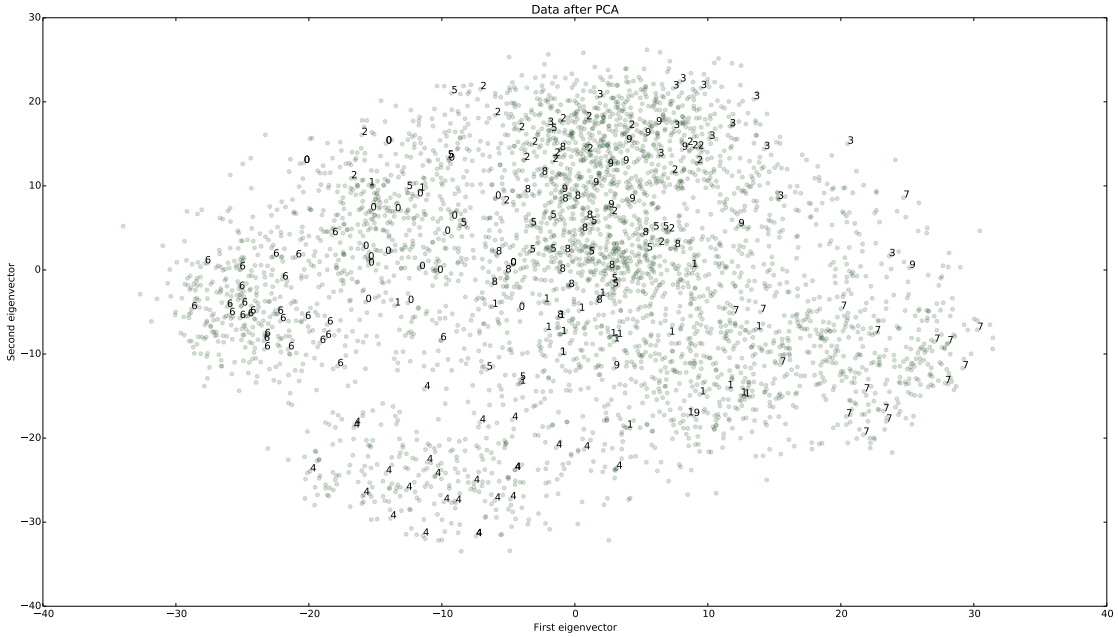

Figure 2: Data plotted in the space of two principal components. Only the label of 200 datapoints are shown.

Submission Policy

- Prepare the report and code. Only electronic submissions through Ninova will be accepted no later than May 3, 10pm
- You may discuss the problems at an abstract level with your classmates, but you should not **share or copy code** from your classmates or from the Internet. You should submit your **own, individual** homework.
- Academic dishonesty, including cheating, plagiarism, and direct copying, is unacceptable.
- Note that your codes and reports will be checked with the plagiarism tools **including previous years submissions**!
- If a question is not clear, please let the teaching assistants know by email **kivrakh@itu.edu.tr**.

Bonus marks (10pts)

- Clarity and nicely described report
- Using Latex template for the report

Deductions (-10pts)

- Spelling errors.
- Messiness
- Lack of content.
- Irrelevant / mistaken content.

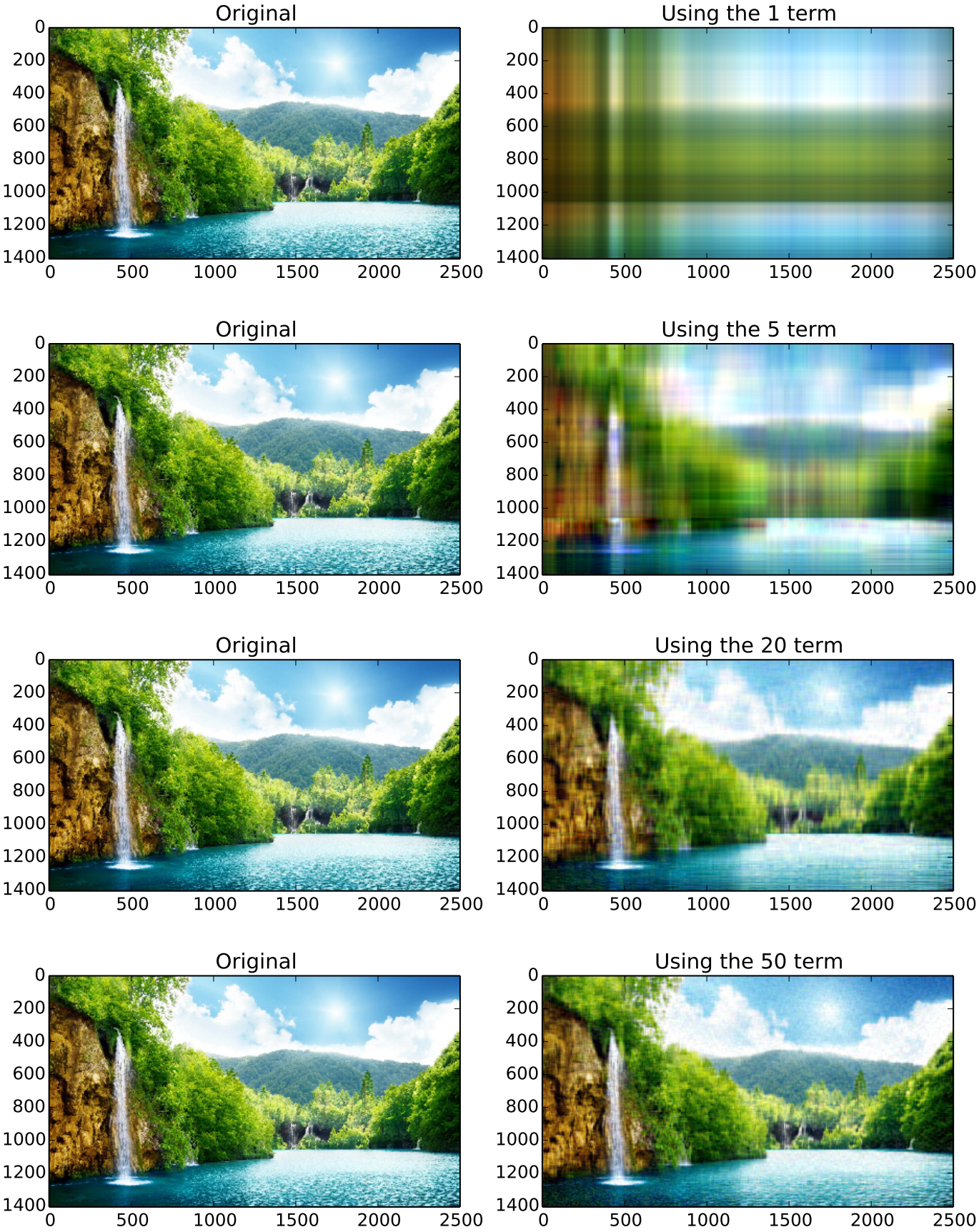

Figure 3: The original image and its the compressed results are displayed.