Linear Algebra Final Project Report

Tianyang Ge

December 28, 2016

Abstract

The final project is a face-recognization code. It uses Principal Component Analysis(PCA) to reduce the dimension:for me, i use the built-in functions which MATLAB provided and it is pity to miss bonus credits. This report will describe the algorithm and more details of the code and related linear algebra. At Last, Thank you for watching my report.

0.1 The algorithm and the code

Considering the algorithm, it mostly consists of PCA and solve $\lim_{x} : \frac{1}{2}|y - Ax|_2^2 + \lambda|x|_1$. Luckily, the project has offered feature_sign. Firstly, how can we search the file to find .pgm? After the user input the number of Trainsamples: n. I use 'for' circle to get all images' path, and then randomly select n images from every files. Imread all images that turn them into matrices. Change their size and make them vectors. Store them into a large matrix, for example, every image turn into a 1000 rows vector and 100 images store together, a 1000rows and 100 columns matrix is created. This is Trainsamples matrix. There is a difference between store trainsamples and test images: trainsamples just become a whole matrix, while test stores in a struct separately. In addition, I resize the matrix to keep the same size. Secondly, PCA function is used to handle the Trainsamples matrix. Now we get scores, coefficient and latent. Scores means new A which based on new basis and coefficient is the new basis. Assume test vector is Y, we change the basis to coefficient and it turns to y. Then, latent is a good tool to reduce high dimensions. I choose 0.95 to select the eigenvalues which left to calculate. The dimensions reduce to very low level about 40 rows. For featuresign function, we need to input scores' and y. The lambda I set to 0.0006 and x_initial is all zeros. At last, we have x. Thirdly, x represents the similarity with basis. I compare the elements and choose that larger than 0.2 to calculate the file it comes from. I choose norm 2 to calculate errors. The accuracy is about 70 percent.

0.2 Related linear algebra

When we talk about linear algebra knowledge, the project covers eigenvalue and eigenvector, basis change. The eigenvalues represent the variation of the row,so when discarding some lines we could highly improve the speed of the code. On the other hand, we need to change Y's basis. $[\varepsilon_1\varepsilon_2\varepsilon_3....]y = \mathbf{I}Y$ The knowledge of linear algebra is not the whole of the project. However, it is important.

0.3 Possible improvements

The speed of code is not satisfying. It costs too much time to run and get the result. Maybe I could simplify the proceed after learning Matlab further. There is a special warning seldom might appear showing that there is a special matrix cannot be handled in the

feature_sign in early versions of the my code. After I improve the code and debug, it disappeared. Maybe it need to be considered and find the cause. In the end, the accuracy is still needed to improve because it certainly could turn to 100 percent.