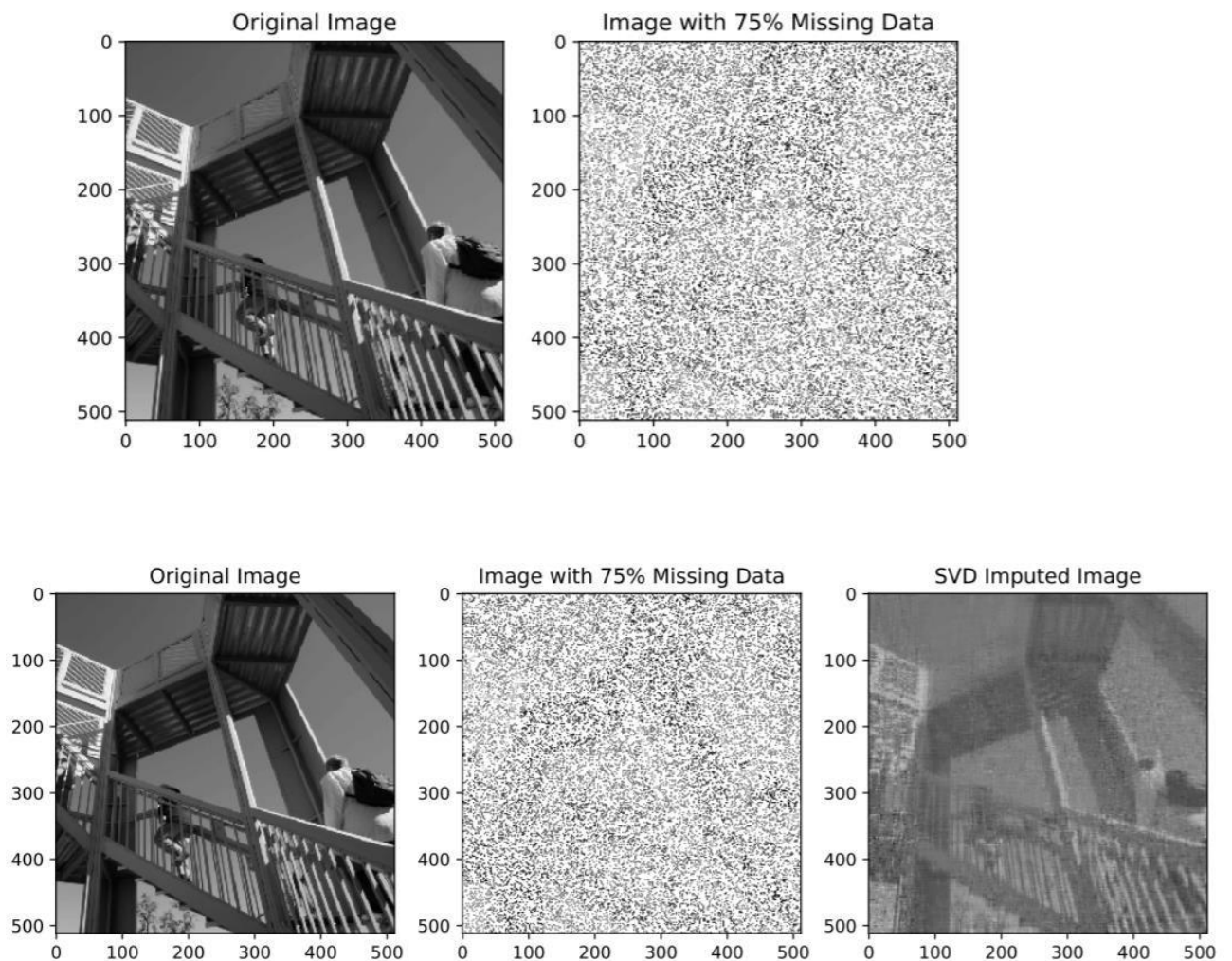# Homework 3: SVD Imputation & Kernel PCA

Zhi Ye

yezh@student.ethz.ch
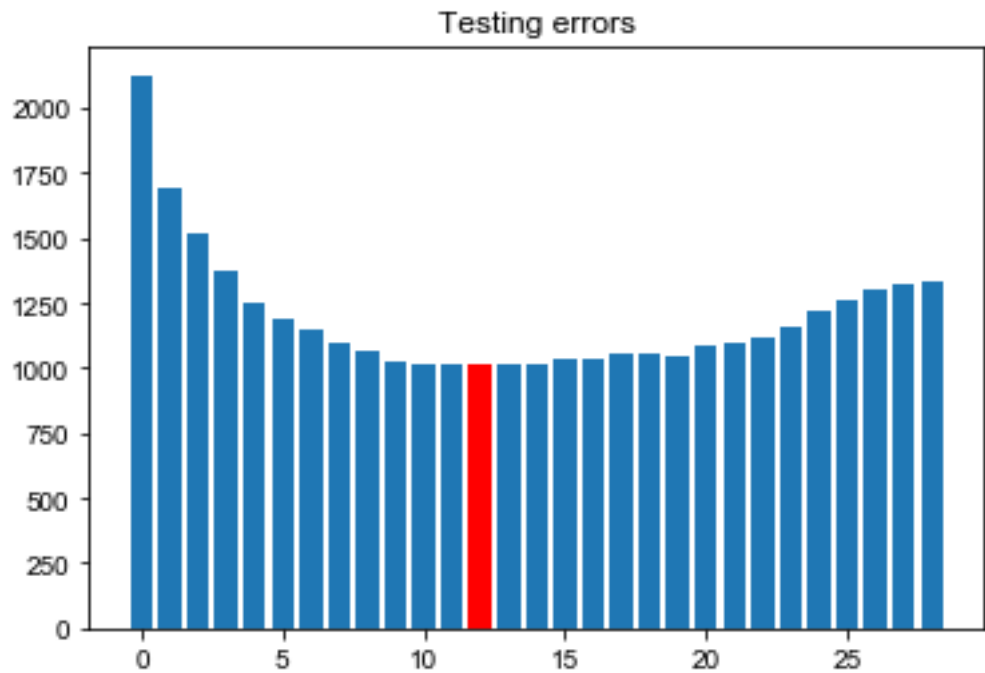
Exercise 1

Exercise 1.a

Exercise 1.b

Optimal rank r is 13, and correspond minimum mean squared error is 1010.54.

Testing errors

Exercise 1.c

The test error will increase dramatically if the rank is too large or too small, and the quality of imputed images will become worse. Hence, the rank should be around the middle to obtain a high quality imputed image and minimum test error.

Exercise 2

Exercise 2.a

It's difficult to separate these two kinds of data after the linear dimensionality reduction into 2-dimension, the PCA is not helpful under this condition.
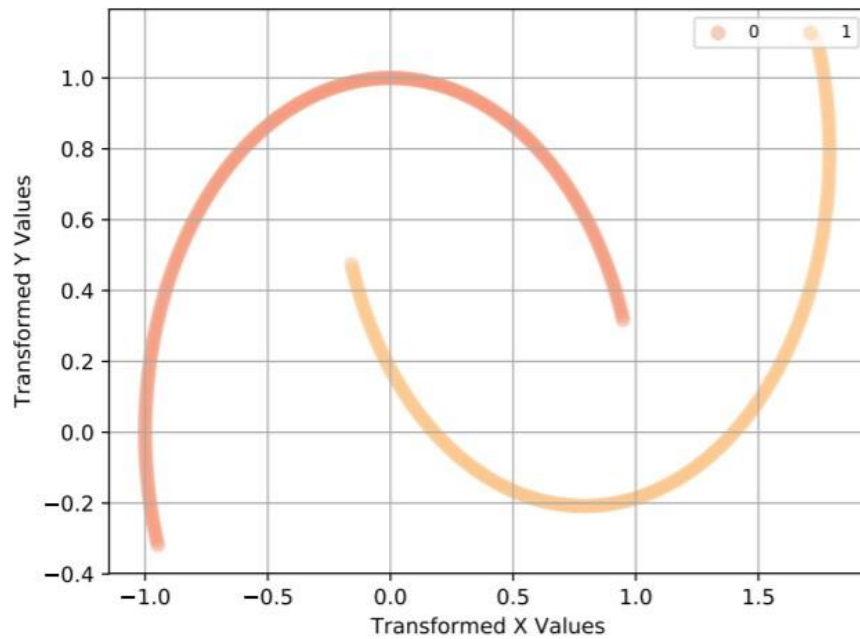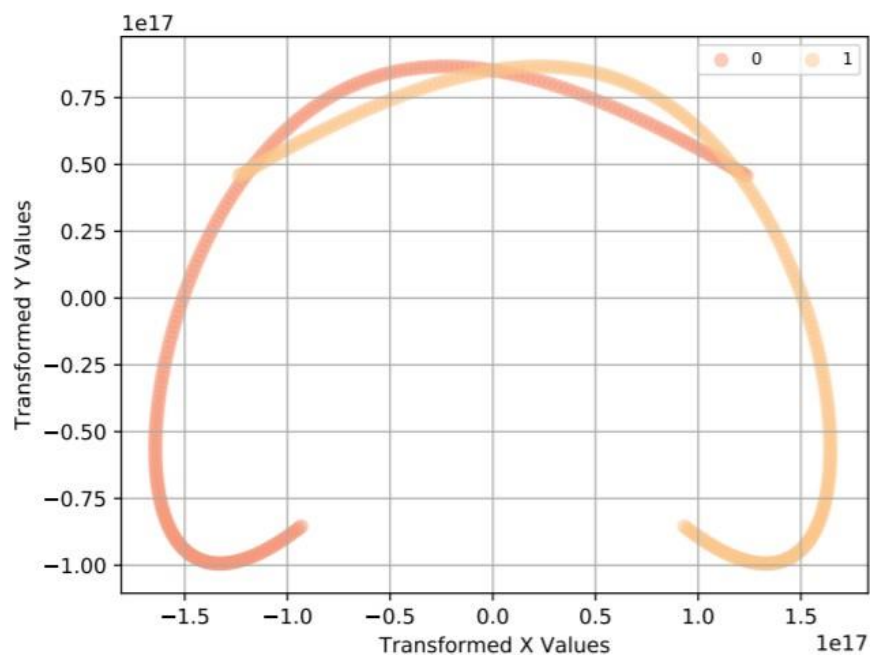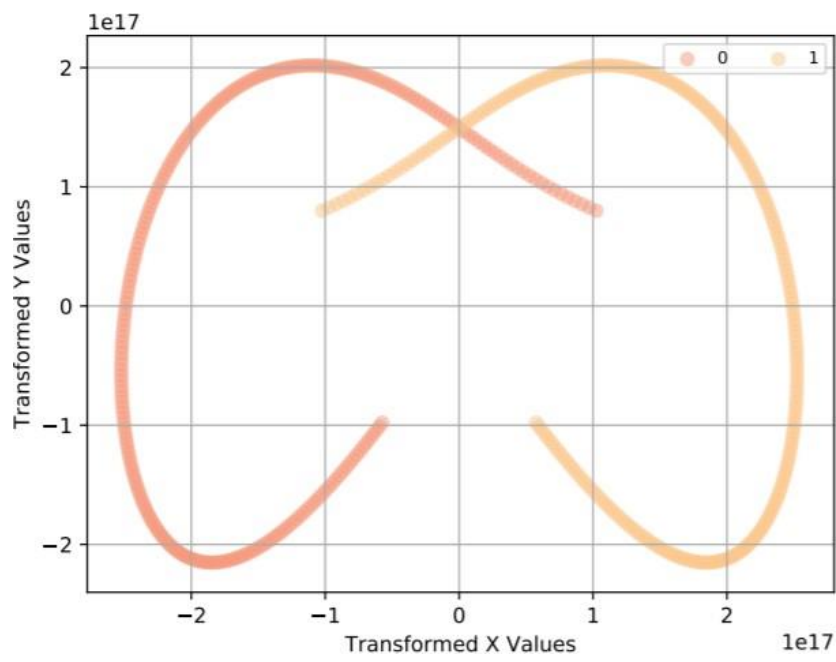
Exercise 2.b

The python files are attached.

Exercise 2.c

According our results, with the increasement of γ, these two kinds of data after the non-linear dimension reduction by using Kernel PCA become easier to separate. But we only test γ to 20, and $\gamma = 20$ is the optimal γ so far.
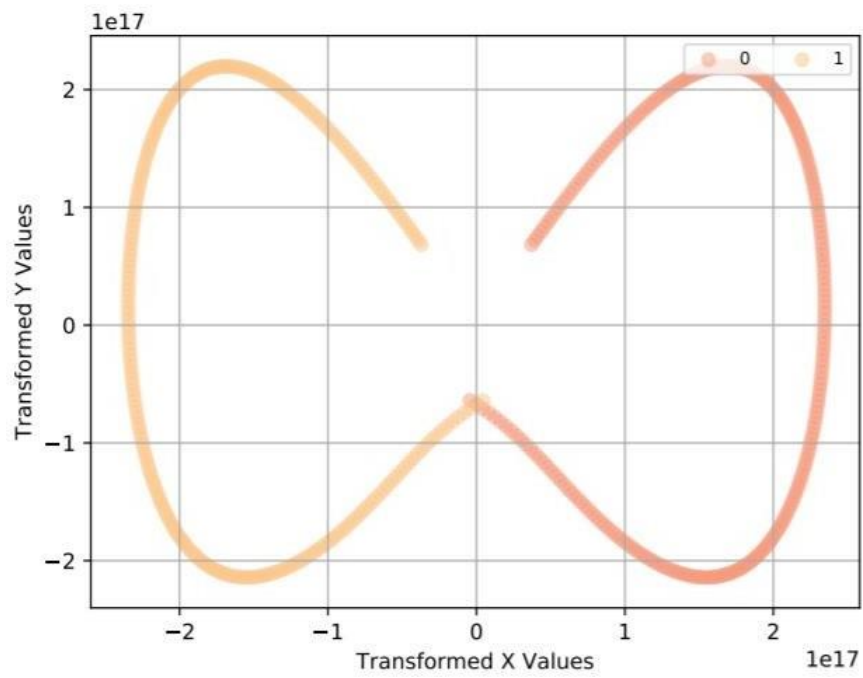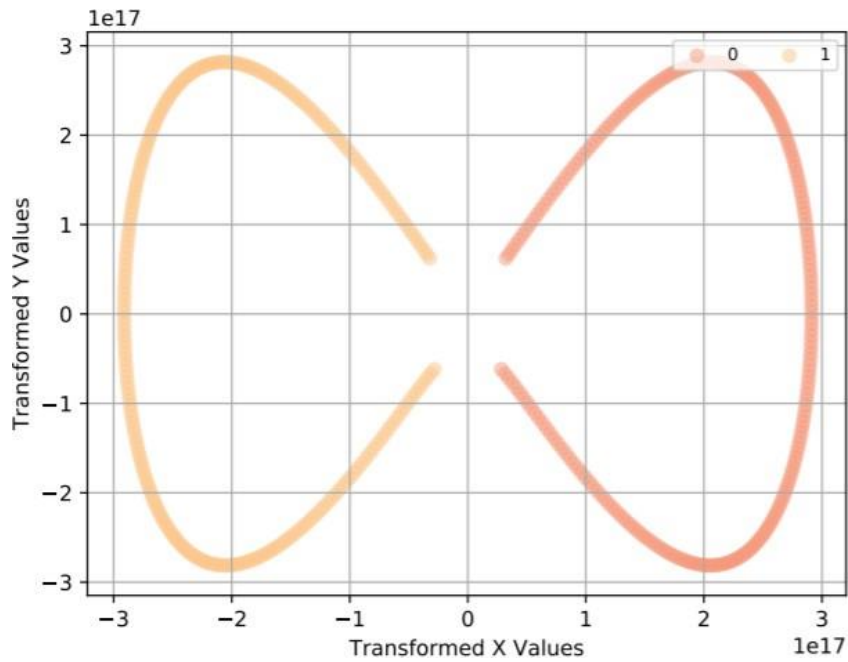
$$\gamma = 1$$

γ = 5

γ = 10

γ = 20

To find the optimal γ, we can use cross-validation or Bootstrap to find a γ that could separate these two kinds of data easily after non-linear dimension reduction. And we also need to choose a metrices to qualify our results' quality, to the good quality of separation means the distance between two kinds of data is the biggest.