Short write-up including the following:

- **Comparison of errors 'in-sample' and 'out-sample' for each value of lambda**

Lambda	0.001	0.01	0.1	0.25	0.5	0.75	1
Ein	0.0230	0.0230	0.0231	0.0233	0.0236	0.0240	0.0239
Eout	0.0236	0.0236	0.0232	0.0232	0.0236	0.0244	0.0244

Table 1: Error rate comparison with lambda

The stability condition of learning rate for a specific model is 0<lambda<1. In theoretically, the error rate increase with learning rate but it is not true for all models. Therefore, choosing best lambda value could be tough. From the above table, we see that error rate (in sample and out sample) has been increased with learning rate.

- **Choice of best value of lambda**

From the table 1, it is shown that the error rate has close relation with the learning rate. For a learning algorithm, we always try to minimize the in sample and out sample error. There are no hard-fast rules to select learning rate for a model. So, empirical method could be best option to select best learning rate. In our case, we pick the best lambda when the error rate is smaller.

- **Discuss effect of lambda on the non-linear transform weights.**

  In our work, we have used 3rd order polynomial feature space. So, the VC dimension for non-linear transform becomes 3*(3+3)/2 +1= 10.

  VC dimension size has raised due to the high order polynomial as well as weight vector dimension.

  The regularization parameter, lambda controls the invertible matrix (identity matrix). The high lambda leads the algorithm convergence fast but increase the error rate. So, the appropriate lambda is calculated using validation.

- **Plot of final classification curve using best lambda and corresponding average weight vector**

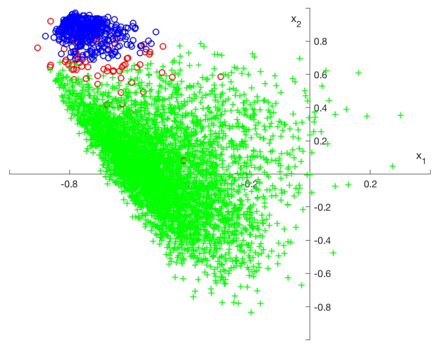

Fig. 1: The figure of final classification of dataset. The red dots represent the misclassified data points. Blue color shows correctly identified 1's value and green present the non 1's value.