Project 1 extra description:

1. After you loaded the dataset, you should have 4 matrices, trX, trY, tsX, tsY. For trX and tsX, each row represents a digit, you can see that their dimension is 28 \* 28 = 784
2. The number of rows represents how many digits there are, so trX should have 6265+5851 rows, tsX should have 1028 + 974 rows.
3. trY and tsY are labels of training set and testing set respectively.
4. For feature extraction, you have to calculate the mean and s.d., of each digit. After the extraction, you should have a new trX and new tsX. Let’s call them trX\_new and tsX\_new, BOTH of them have only two columns, instead of 784.
5. You should use trX\_new and tsX\_new for BOTH naive bayes and logistic regression.

For Naive Bayes:

1. First separate trX\_new into two sets, one set contains only digit 7, another contains only digit 8.
2. Calculate the mean and covariance matrix, for each of these sets.
3. The covariance matrix should be a diagonal matrix, as you are doing naive bayes, the covariance of the two features should be independent.
4. Use them to calculate the required probability.

For logistic regression:

1. Just that the boundary by using gradient.

For the submission, you should upload a report in .pdf and at least one .py file (or .m if you are using matlab)

For the report, it contains several parts. The following is an example from a student last semester:





