

Q2b

As L1 regularization produces a sparse solution, hence doing feature selection. The less important coefficient will have their coefficients equal to zero and hence the more important features will be left.

Q3 c

3 and 8 are the most confused digits, as 8 was confused as 3 for 7 times. The easiest to predict was 2.

Q3 d

Points in the plot refer to be the projection of the data points of the image towards the higher variance direction.