Part A Report

Q1- I assumed p(y) was just how much it occurred in over how many times all classes occurred, as for the p(x|y), we derived it from Bayes' equation, accounting for the mahalnobis distance instead of Euclidean to account for the "shape" of the data or the "covariaine" matrix.

Q2- I didn't estimate those parameters, rather I calculated them, as mentioned in the question above, as for the mean and covariance, I just took the training labelled data set and took the average of the features of each class to create the mean vector for each class, and as for the covariance matrix, I just measured the "deviation" of each picture from mean of each class, and then squared it and took the "expectation" or "average" to get the variance and covariance matrix.

Q3- We need to regularize, in case there is a constant feature in the images, like for example, if there's a constant black pixel in the corner, the variance of that feature will be zero, which will cause the matrix to become non-invertible and stall our calculations, that's why we assume we add a very small value all of the features to have an invertible matrix.

Q2-1:

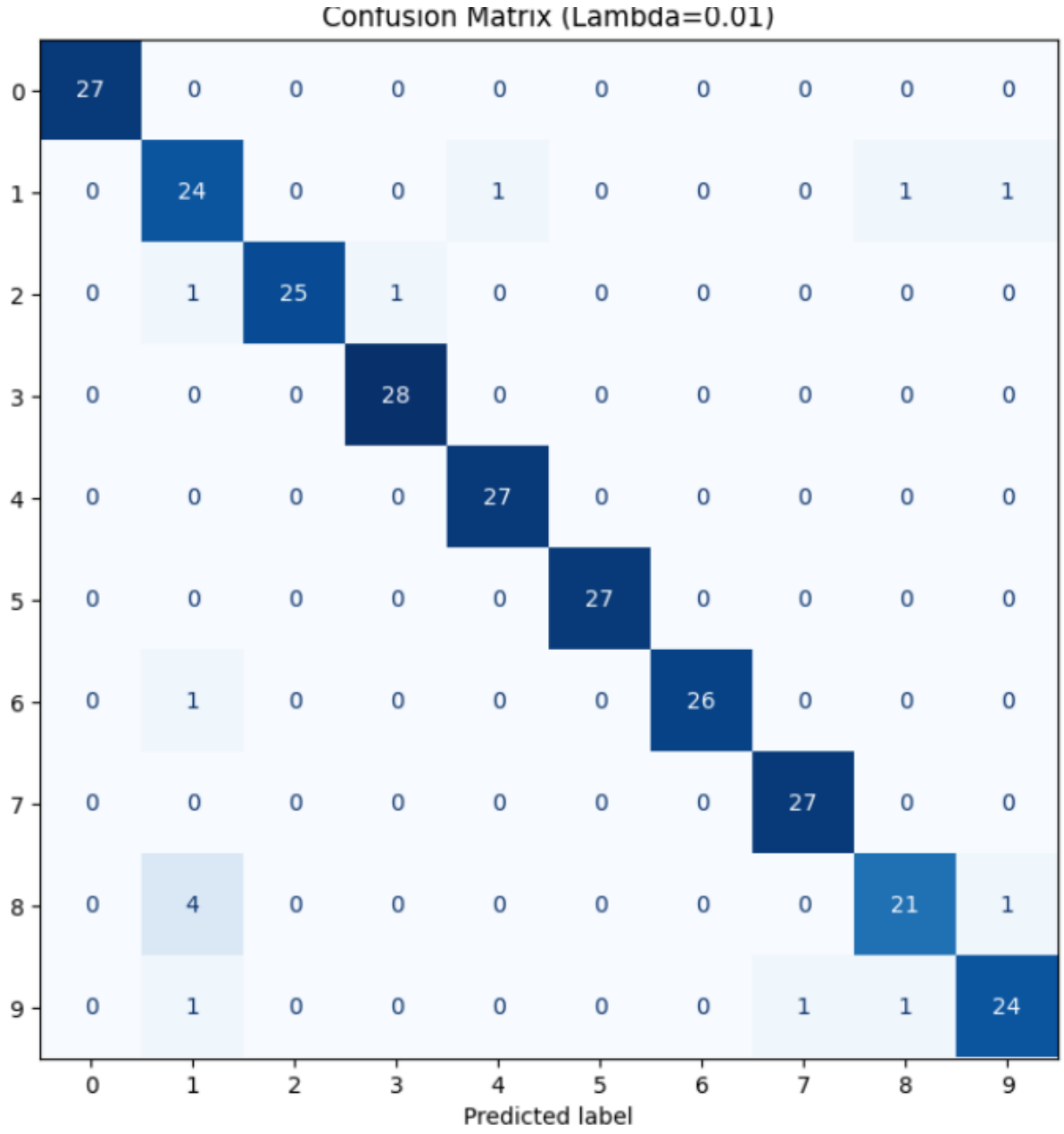
Lambda	Val Accuracy
0.0001	0.9481
0.001	0.9481
0.01	0.9519
0.1	0.9444

Best Lambda found: 0.01

Q3:

Classification Report (Macro-averaged):

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	27
1	0.7742	0.8889	0.8276	27
2	1.0000	0.9259	0.9615	27
3	0.9655	1.0000	0.9825	28
4	0.9643	1.0000	0.9818	27
5	1.0000	1.0000	1.0000	27
6	1.0000	0.9630	0.9811	27
7	0.9643	1.0000	0.9818	27
8	0.9130	0.8077	0.8571	26
9	0.9231	0.8889	0.9057	27
accuracy			0.9481	270
macro avg	0.9504	0.9474	0.9479	270
weighted avg	0.9506	0.9481	0.9484	270



Q4-
From the confusion matrix, the number eight was wrongly classified the most times, as “1”.

Q4-2:
Testing different values for lambda, we got an increase in accuracy of ~1%

Q4-3

Strengths: it captures the the relationship between the different features and takes their “shape” into consideration, or in other words, the variance of each feature into consideration.
Since a small change in feature, x1, could be more significant than a big change in x2 for e.g.