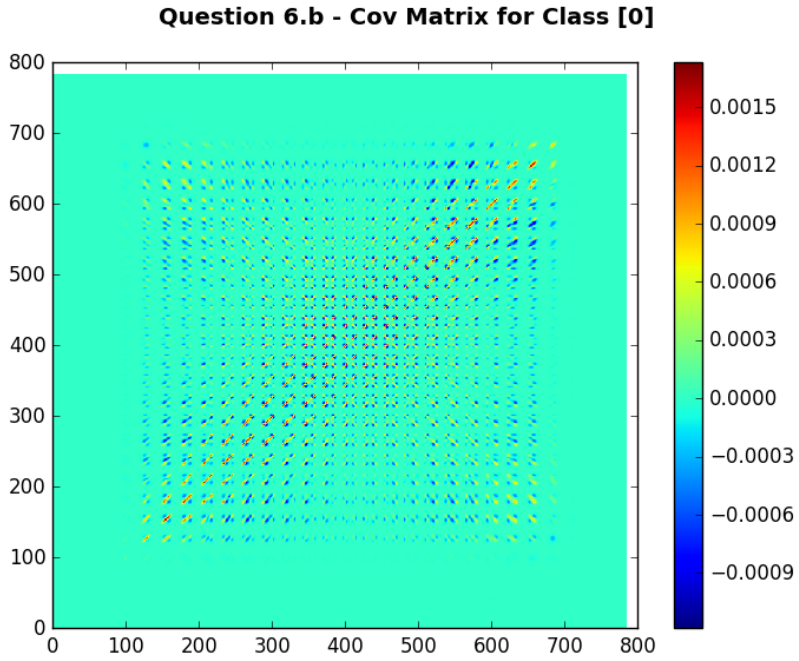
Lev Golod, 3032125890, CS 289 HW3, Question 6

Part B

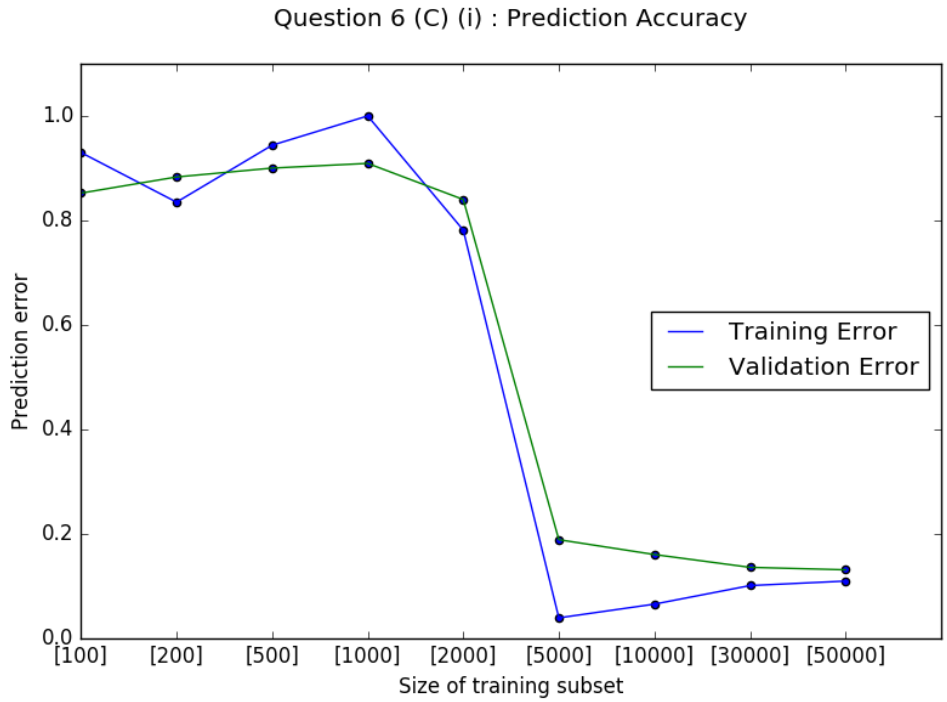


In this plot we observe that the diagonal elements (the variances of each pixel) are mostly positive. However, most of the off-diagonal terms (the covariances of pixels with other pixels) are mostly close to zero, or even slightly negative. This means that within a class, our features are not strongly correlated with one another. This is good. If we had too many strong covariances (large off-diagonal) terms, it would either give us a singular matrix (not invertible) or a nearly-singular matrix (inversion is unstable). Either of these would make fitting a Gaussian model more difficult.

APENDIX: Code

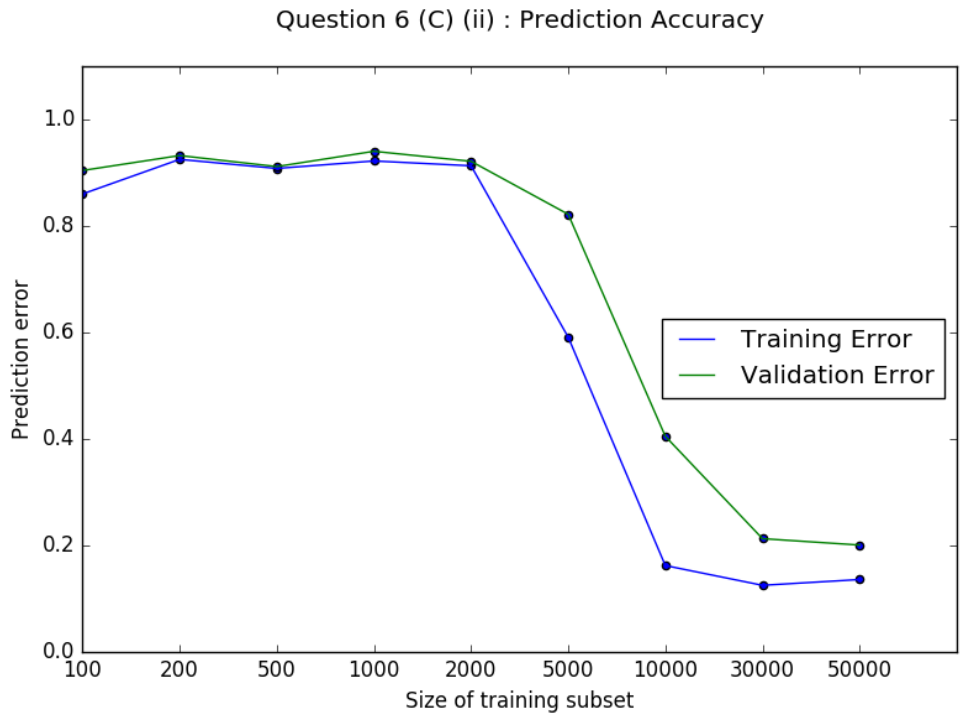
```
fig = pcolor(mle_cov[0])
colorbar()
plt.suptitle('Question 6.b - Cov Matrix for Class [0]', fontsize=14, fontweight='bold')
plt.savefig('./plot/q6_b.png', bbox_inches='tight')
```

Part C (i)



# Subset size - **LDA**	Training Error	Validation Error
100	0.93	0.85
200	0.83	0.88
500	0.94	0.9
1000	1	0.91
2000	0.78	0.84
5000	0.04	0.19
10000	0.07	0.16
30000	0.1	0.14
50000	0.11	0.13

Part C (ii)



# Subset size - **QDA**	Training Error	Validation Error
100	0.86	0.9
200	0.93	0.93
500	0.91	0.91
1000	0.92	0.94
2000	0.91	0.92
5000	0.59	0.82
10000	0.16	0.41
30000	0.13	0.21
50000	0.14	0.2

Lev Golod, 3032125890, CS 289 HW3, Question 6

Part C (iii)

I achieved about 20% validation error with QDA, and 13% with LDA. LDA worked better for me. I think this is because of the greater stability involved in calculating one common pooled-within-class covariance matrix for LDA, as opposed to 10 different class-specific covariance matrices with QDA. With LDA, you have one data matrix with N rows (observations) and p columns (features). With QDA, on average, each covariance matrix has n = N/10 rows and p columns. As discussed in question 5, having p >= n is one way for a covariance matrix to be non-invertible, requiring some sort of modification to calculate the inverse. This was more likely to happen with QDA, leading to lower-quality estimation of the underlying Gaussian distributions, and lower-accuracy predictions.

Part D

I added one new feature. This feature took a value of 1 if all pre-existing features for that observation were 0. Otherwise it took value 0.

APPENDIX: Kaggle scores
My username on Kaggle is LevGolod1

Submission	Files	Public Score	Selected?
Thu, 23 Feb 2017 06:05:51 Edit description	spampred.csv	0.77220	☐

Submission	Files	Public Score	Selected?
Thu, 23 Feb 2017 05:07:03 Edit description	mnistpred.csv	0.88360	☐