**CSC411 Homework 5**

**Question 1: Gaussian Discriminant Analysis**

**Part A, B: Average Conditional Log Likelihood, Accuracy**

Output of program:

Part A: Average Conditional Log-Likelihoods

Train: -0.12462443666863048

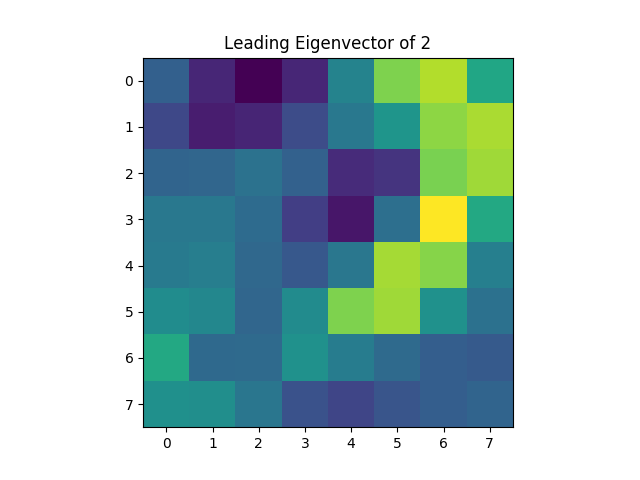
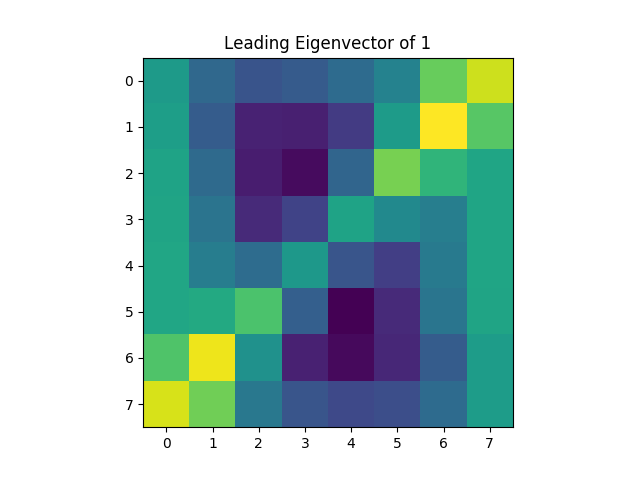
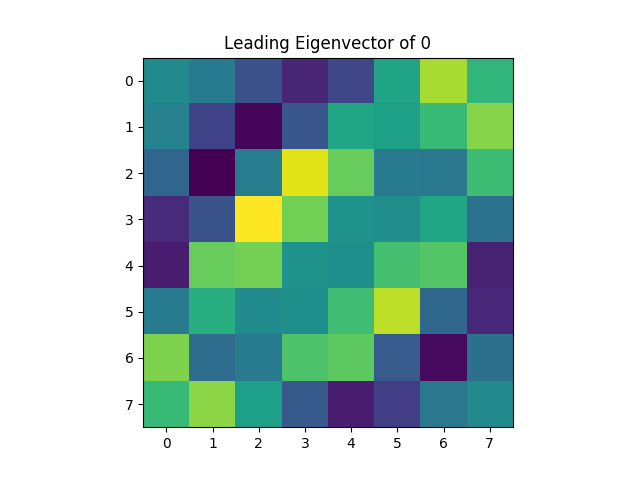
Test: -0.19667320325525584

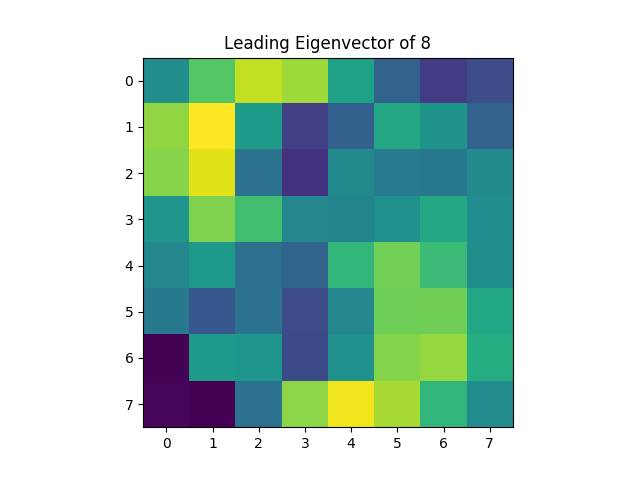
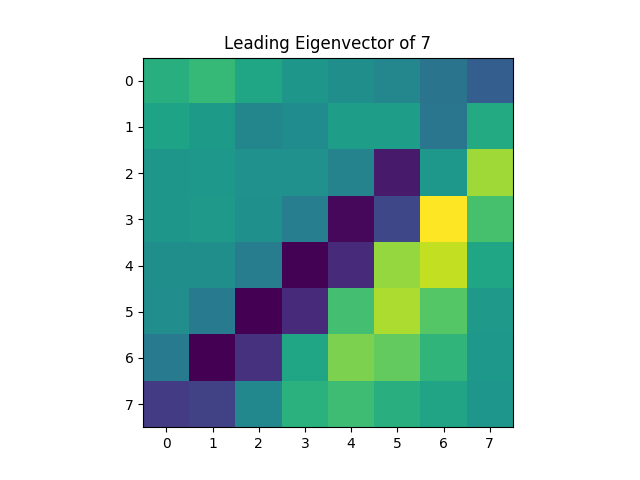
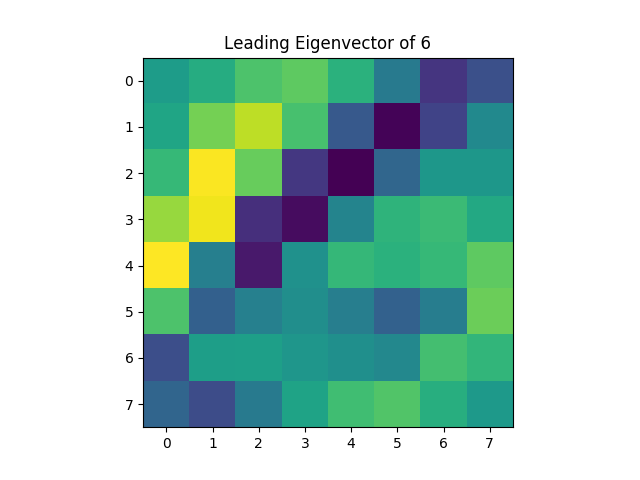
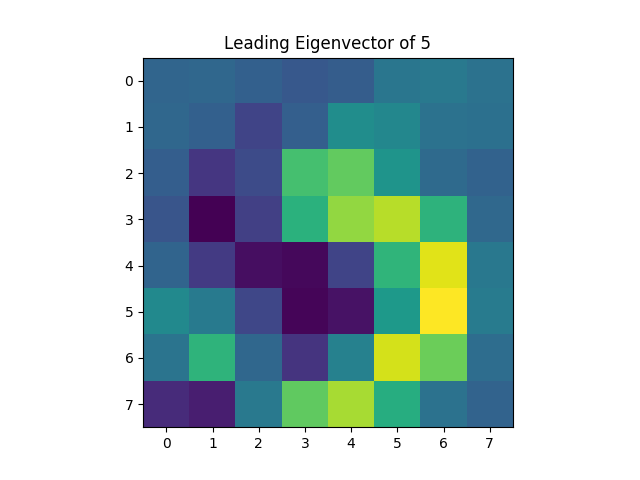
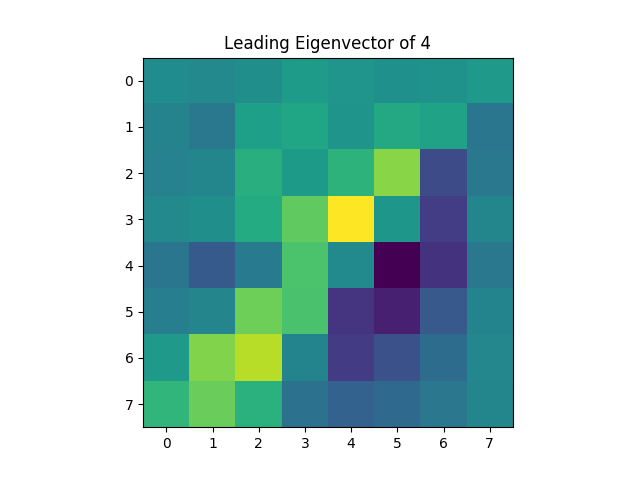
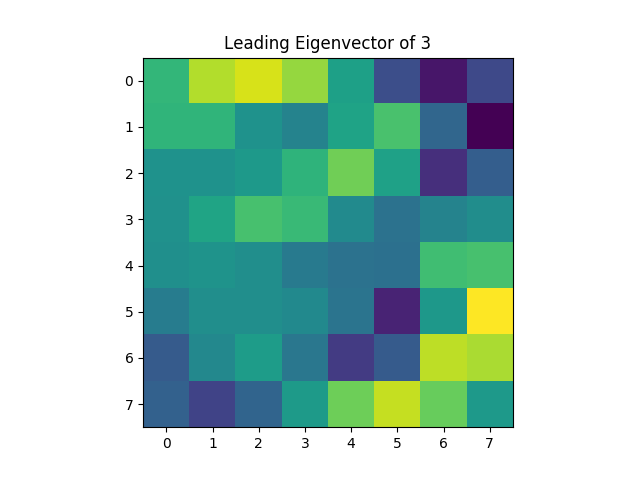
Part B: Accuracy

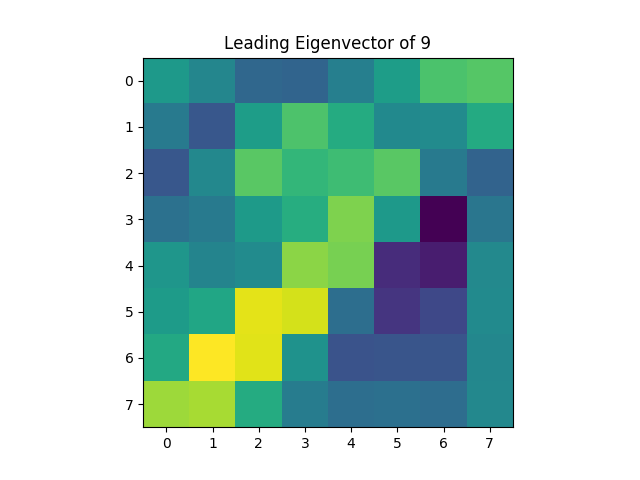
Train Accuracy: 0.9814285714285714

Test Accuracy: 0.97275

**Part C: Visualizing Leading Covariance Eigenvectors**







**Question 2**

**Part A: Posterior Distribution**

Bayes rule:

Dirichlet as prior:

Likelihood of data given

Note that is a one-hot vector. This means that all elements of except for the element corresponding to the correct class will “cancel out” of the product. If class j has occurences, then the total power of will be

For the purposes of this question, we can reason about in proportional terms:

Merging the two products together:

Note that the above proportionality means that the conditional probability also obeys a Dirichlet distribution

Posterior predictive probability

Let

Probability of must be integrated over all possible .

Probability of is equal to

Integration over means the above is an expectation of

Recalling expectation of Dirichlet distribution

N is the sum of all

**Part B: MAP of**

Recall proportionality of :

Since is directly proportional to the product of powers of theta, we need only maximize the product

Logarithm function is monotonic, so maximizing the logarithm is also equivalent to maximizing the input of the logarithm

Applying the logarithm to the product

Note that , so Lagrange multipliers must be used for maximization

Taking derivatives of the Lagrangian equation:

Setting derivatives to zero to find maximum

Combining these equalities, we get:

Substituting back into the definition of