109550069 林宇柔

ML HW2

Part 1.

1. Mean vectors

```
mean vector of class 1: [ 0.99253136 -0.99115481] mean vector of class 2: [-0.9888012   1.00522778]
```

2. Within-class scatter matrix

```
Within-class scatter matrix SW: [[ 2.31560974 -0.95793362]
 [-0.95793362  1.50940428]]
```

3. Between-class scatter matrix

```
Between-class scatter matrix SB: [[ 3.92567873 -3.95549783]
 [-3.95549783  3.98554344]]
```

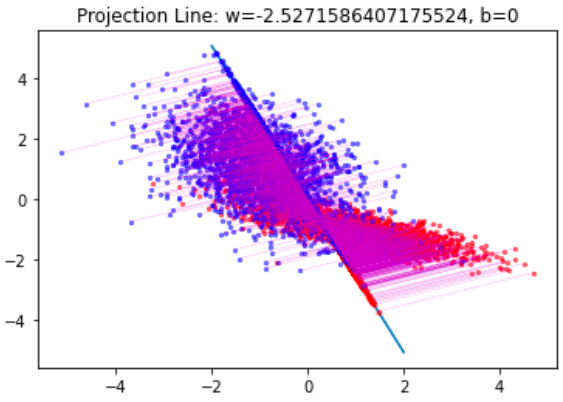
4. Fisher's linear discriminant

```
Fisher's linear discriminant: [-0.36794232  0.92984862]
```

5. Accuracy score on test data with K values from 1 to 5

```
k = 1:
Accuracy of test-set 0.8584
k = 2:
Accuracy of test-set 0.8656
k = 3:
Accuracy of test-set 0.8816
k = 4:
Accuracy of test-set 0.888
k = 5:
Accuracy of test-set 0.8928
```

6. Plotting



Part 2.

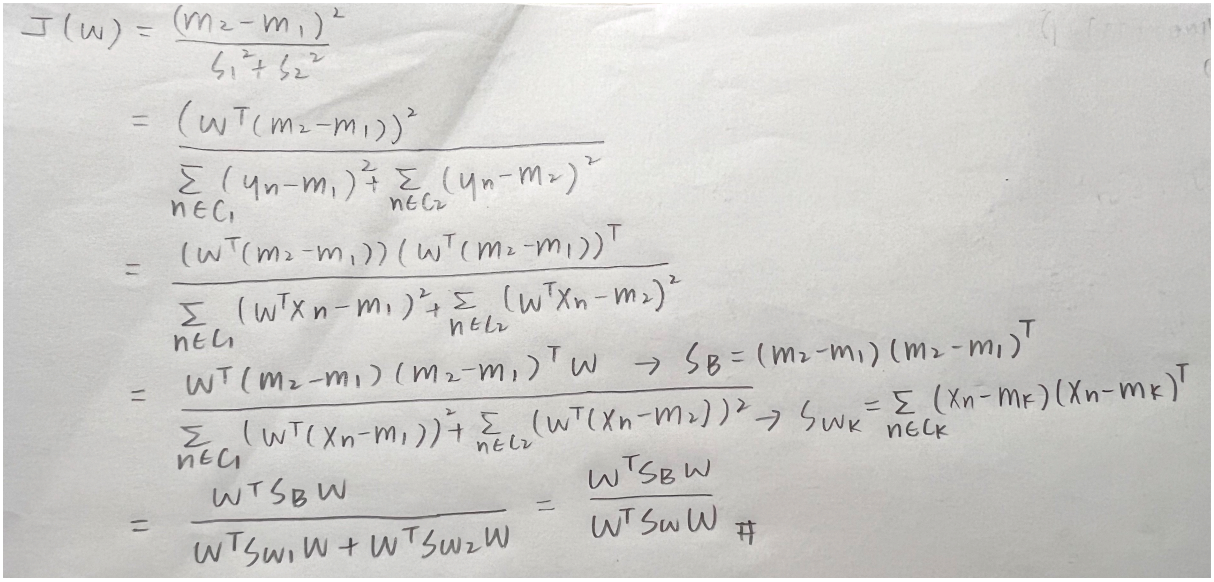
What's the difference between the Principle Component Analysis and Fisher's Linear Discriminant?

Principle Component Analysis focuses on capturing the direction of maximum variation in the data set. Fisher's Linear Discriminant focuses on funding a feature subspace that maximizes the separability between the groups.

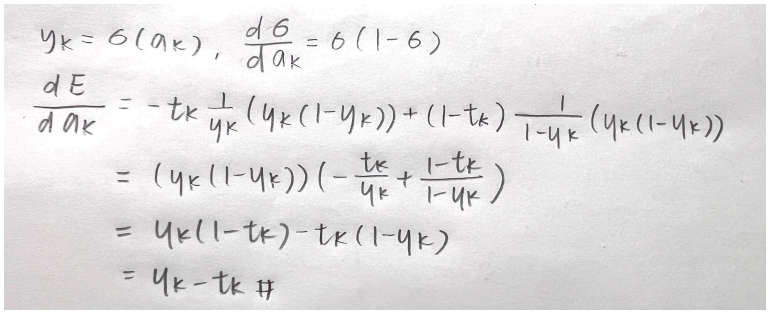
Please explain in detail how to extend the 2-class FLD into multi-class FLD (the number of classes is greater than two).

When it there are only two classes, we add up the two covariance matrices, to extend it to a K-class FLD, the within-class covariance matrix is the sum of all K covariance matrices. As for the between-class covariance matrix, instead of subtracting two means, we get the sum of differences of the mean of each class and a mean in general, also, the rank of this matrix would be at most K-1. The optimal w is the eigenvector of the inverse of the within-class matrix times the between-class matrix, which corresponds to the largest eigenvalue.

By making use of Eq (1) ~ Eq (5), show that the Fisher criterion Eq (6) can be written in the form Eq (7).



Show the derivative of the error function Eq (8) with respect to the activation ak for an output unit having a logistic sigmoid activation function satisfies Eq (9).



Show that maximizing likelihood for a multi-class neural network model in which the network outputs have the interpretation yk(x, w) = p(tk=1 | x) is equivalent to the minimization of the cross-entropy error function Eq (10).