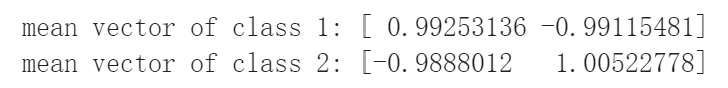
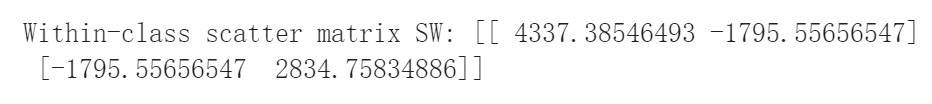
**Part. 1, Coding (60%)**:

1. (5%) Compute the mean vectors (i=1, 2) of each 2 classes on **training data**



1. (5%) Compute the within-class scatter matrix on **training data**



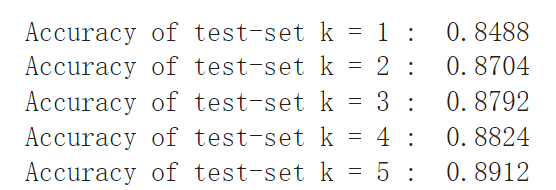
1. (5%) Compute the between-class scatter matrix on **training data**



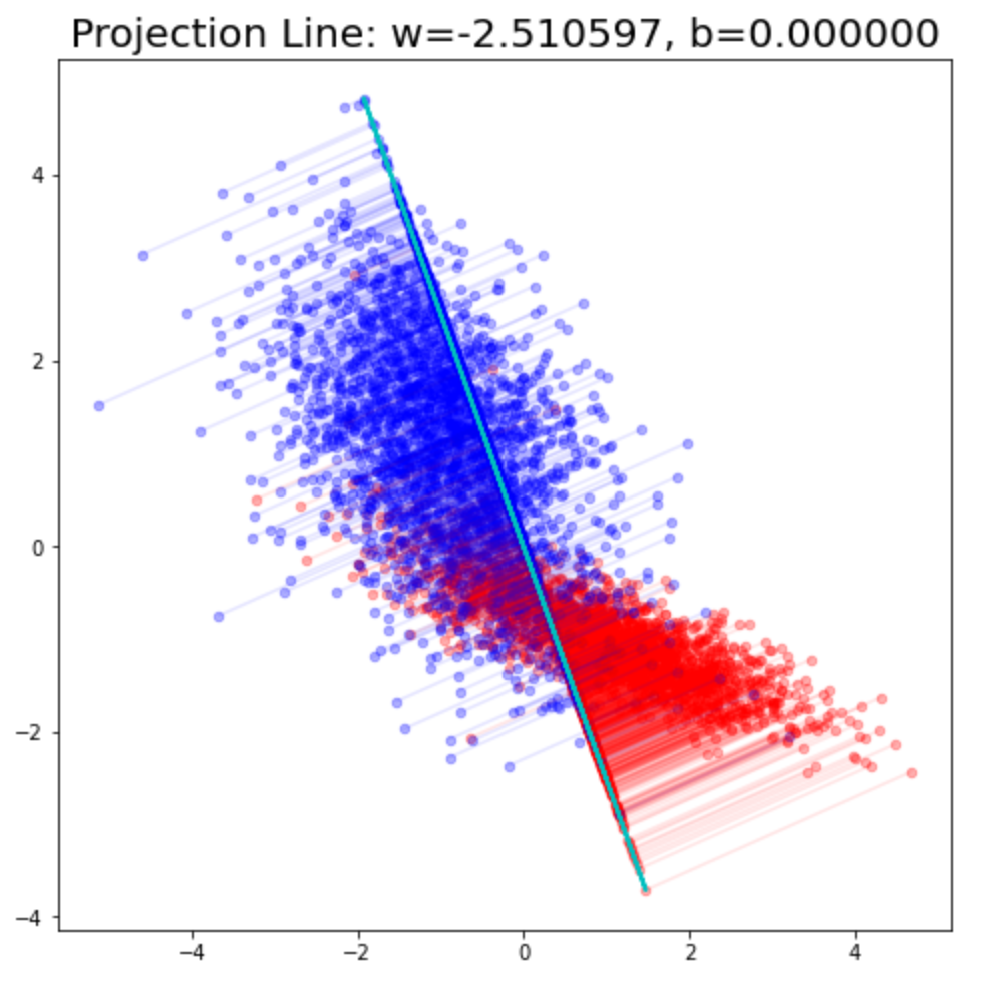
1. (5%) Compute the Fisher’s linear discriminant on **training data**



1. (20%) Project the **testing data** by Fisher’s linear discriminant to get the class prediction by K-Nearest-Neighbor rule and report the accuracy score on **testing data** with K values from 1 to 5 (you should get accuracy over **0.88**)



1. (20%) Plot the **1)** **best** **projection line** on the **training data** and show the slope and intercept on the title *(you can choose any value of* ***intercept*** *for better visualization)*   
   **2)** **colorize the data** with each class **3)** project all data points on your projection line.



**Part. 2, Questions (40%):**

(10%) 1. What's the difference between the Principle Component Analysis and Fisher’s Linear Discriminant?

Principle Component Analysis (shorten as PCA) is an unsupervised dimentionality reduction technique which ignores class labels. PCA projects the higher dimensional data into lower dimension while preserving the significant information.

While Fisher’s Linear Discriminant (shorten as FLD) is a supervised dimentionality reduction technique that acheives data classification simutaneously.  FLD maximize the separation among classes while minimize the separation between classes.

Both of them looks for linear combinations of the features that explain the data best.

PCA is efficient in terms of detection of faults while FLD is optimal in terms of diagnosing the faults.

(10%) 2. Please explain in detail how to extend the 2-class FLD into multi-class FLD (the number of classes is greater than two).

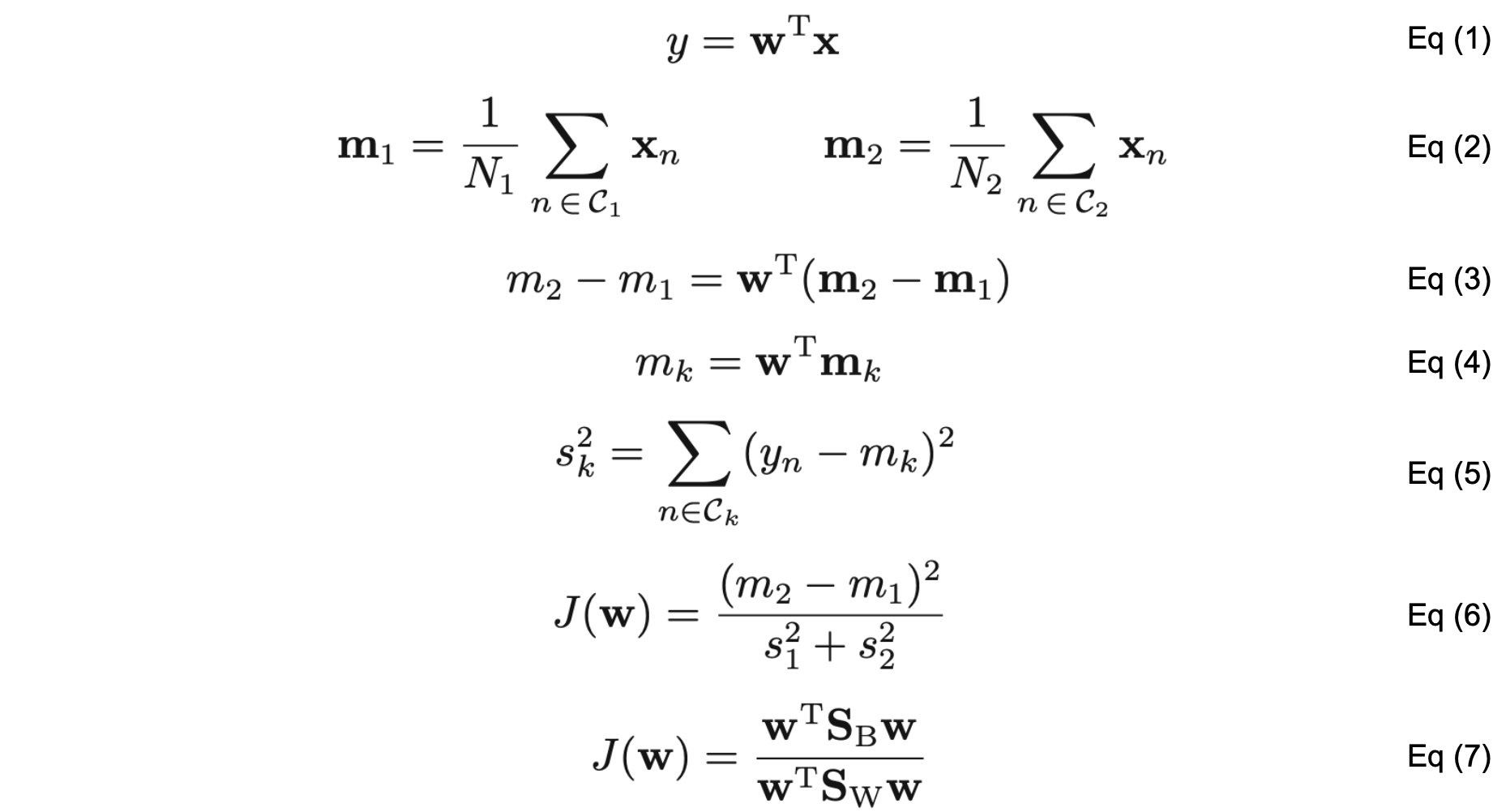
To extend 2-class FLD into multi-class FLD, assume that there are N classes instead of 2, the i-th class is denoted as .

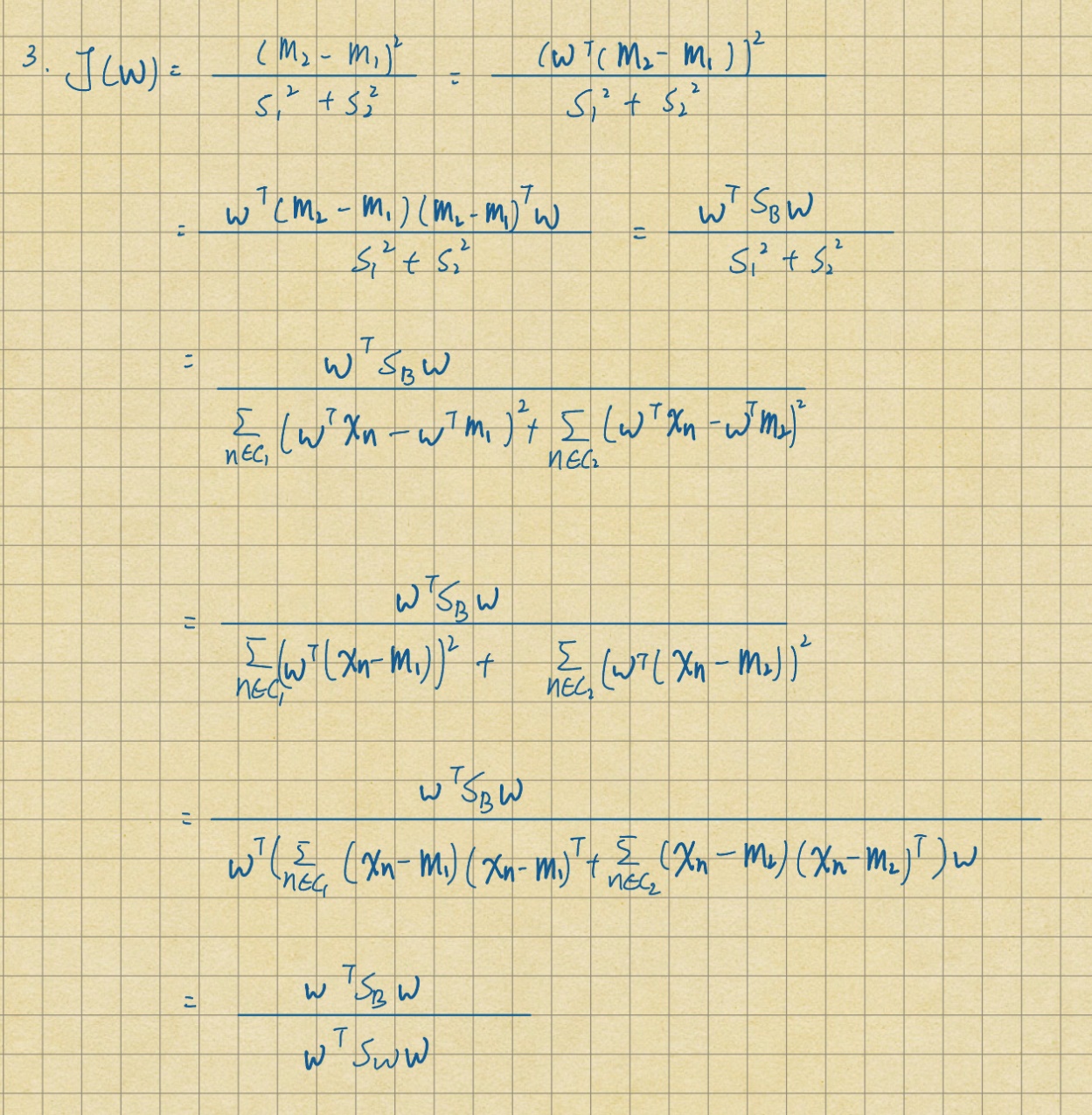
Then we can redefine as , where , and , is the mean of each class i.

The objective from 2-class cannot extend to multi-class directly, so let the objective be . The columns of the optimal are the eigenvectors of that correspond to the largest eigenvalues.

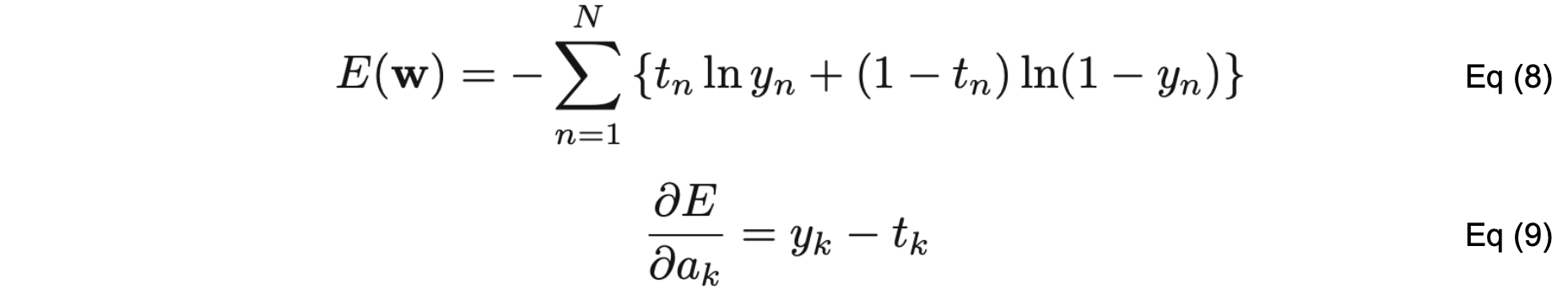
With the formulas above, we can find the projection of the multi-class FLD.

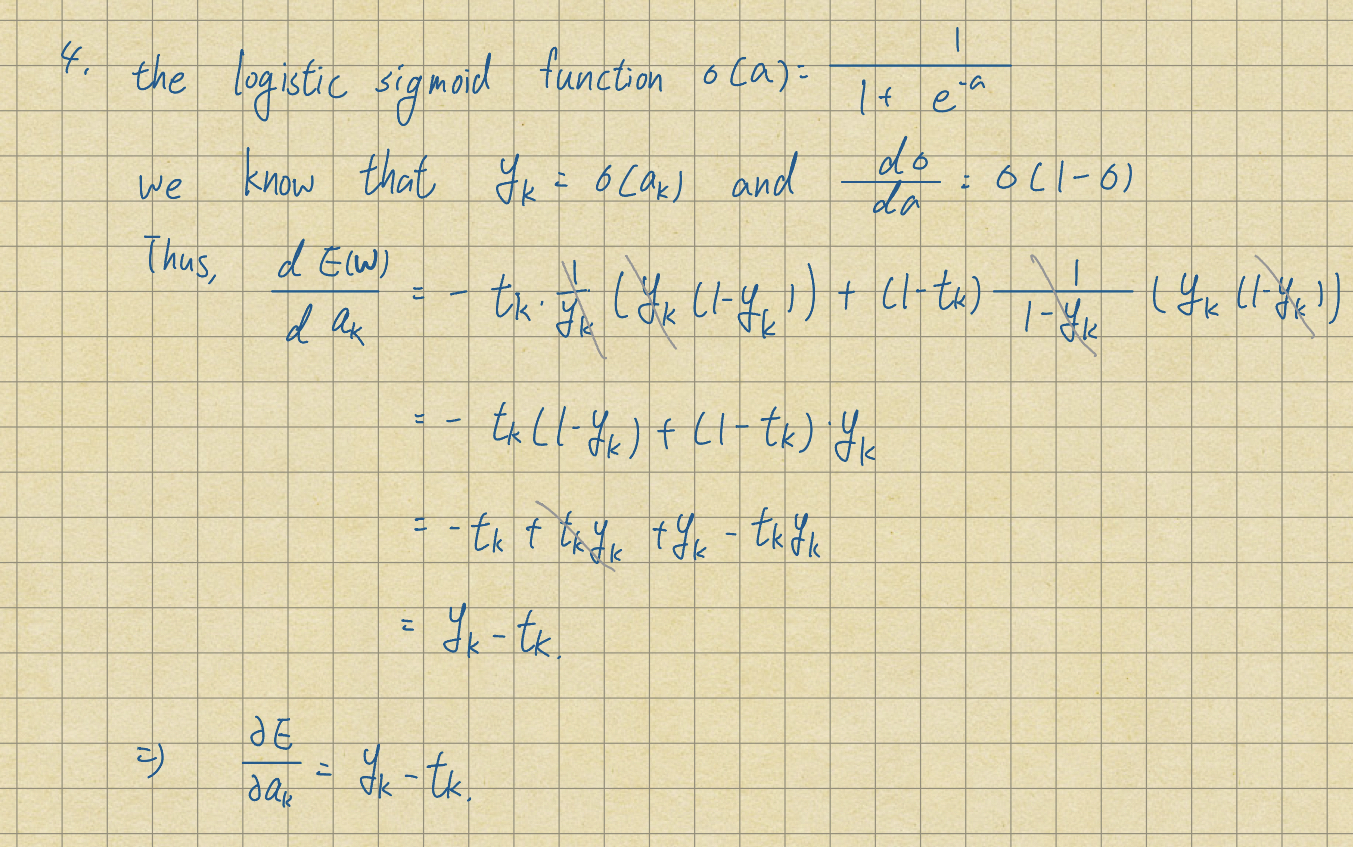
(6%) 3. By making use of Eq (1) ~ Eq (5), show that the Fisher criterion Eq (6) can be written in the form Eq (7).

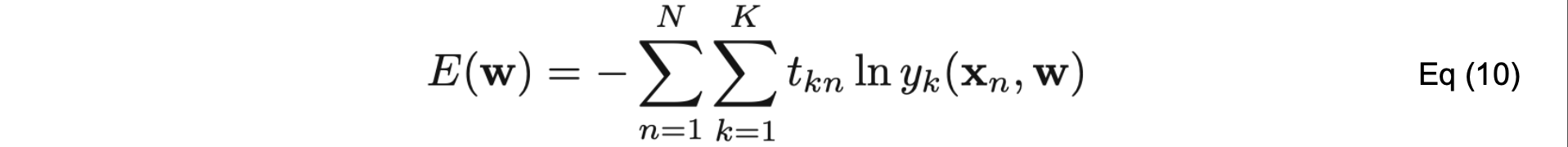




(7%) 4. Show the derivative of the error function Eq (8) with respect to the activation for an output unit having a logistic sigmoid activation function satisfies Eq (9).





(7%) 5. Show that maximizing likelihood for a multiclass neural network model in which the network outputs have the interpretation is equivalent to the minimization of the cross-entropy error function Eq (10).

