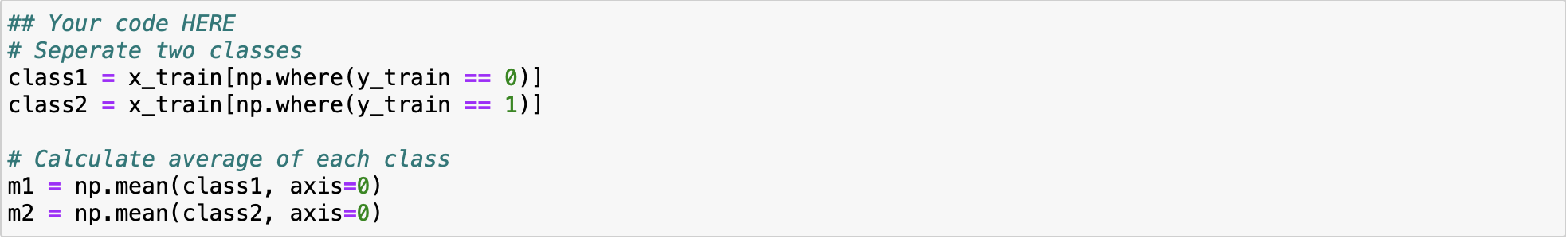
**NCTU Pattern Recognition Homework 2**

M093781 趙宇涵

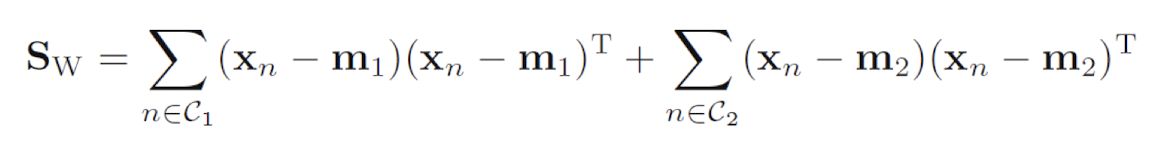
Part 1. Coding (60%)

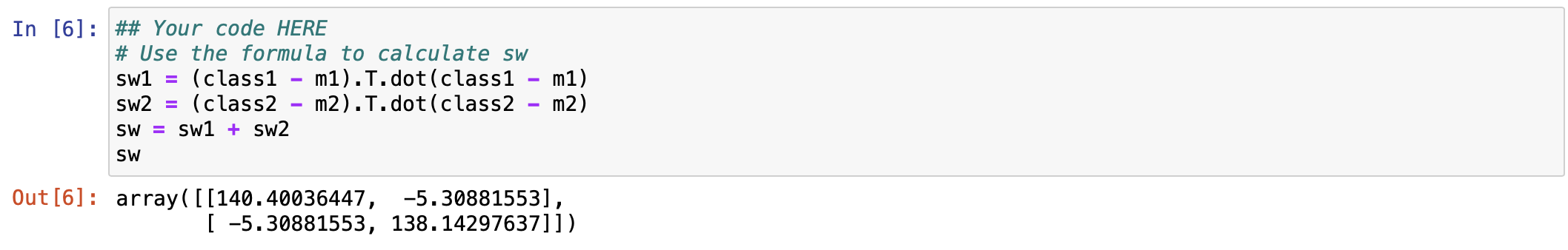
1. (5%) Compute the mean vectors mi (i=1, 2) of each 2 classes on **training data**

First, separate 2 classes. I use numpy.mean to calculate m1, m2.



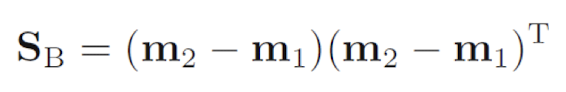
1. (5%) Compute the within-class scatter matrix Sw on **training data**

Use the formula to calculate Sw.



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

Use the formula to calculate SB.



一張含有 文字 的圖片

自動產生的描述

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

*W* is proportional to Sw-1(m1-m2), use norm to get unit vector.

一張含有 文字 的圖片

自動產生的描述

一張含有 文字 的圖片

自動產生的描述

1. (20%) Project the **testing data** by Fisher’s linear discriminant to get the class prediction by nearest-neighbor rule and calculate your accuracy score on **testing data**  (you should get accuracy over 0.9)

Step 1. Define the Euclidean distance by .

Step 2. Define get neighbor function. Sort the all training point by distance, and get the nearest n points’ index.

Step 3. Define predict classification. Based on the nearest points index, find their correlated label by y\_train. Because the n is an odd number, there must be a label more than another. Choose the major label to be the predicted class.

Step 4. Use a for loop to find all predicted classes.

Step 5. Use accuracy\_score function to calculate the accuracy.

Noted. I tried n form 3 to 7, and I can get the highest accuracy 0.912 at n = 5. Thus, I choose n = 5 to be my final answer.

一張含有 文字 的圖片

自動產生的描述

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. Your result should look like the below image (This image is for reference, not the answer)