Mid-Term: Question 2, Part 2

Ian Dover

March 2023

1 Part 2

We know that:

$$\sigma_w^2 = w_0\sigma_0^2 + w_1\sigma_1^2$$

Using the relations:

$$w_0\mu_0 + w_1\mu_1 = \mu_T$$

and

$$w_0 + w_1 = 1$$

and

$$\sigma_T^2 = \sum_{i=1}^{L}(i - \mu_T)^2 p_i$$

we can show that

$$\sigma_B^2 = w_0(\mu_0 - \mu_T)^2 + w_1(\mu_1 - \mu_T)^2$$

can be reduced to the relation which we are trying to prove:

$$\sigma_B^2 = w_0 w_1(\mu_1 - \mu_2)$$

This shows that the clusters which minimize the weighted variances of each also maximize the squared difference in means between the means of each cluster and the population. From there, we can reduce the equation to the equation in which we are trying to prove.