ENGR421

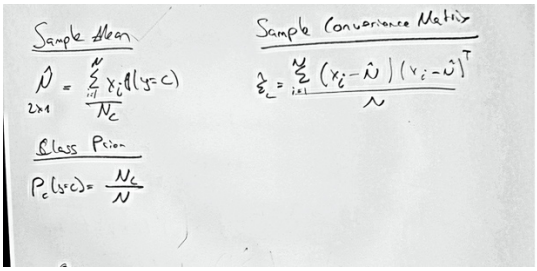
Homework 1 Report
Ata Sayın, 64437

Data Generation

With the giving parameters, class means, covariance matrixes and class sizes are written for multivariate normal density. Points are stacked in order for a total dataset while labeling them according to their class sizes.

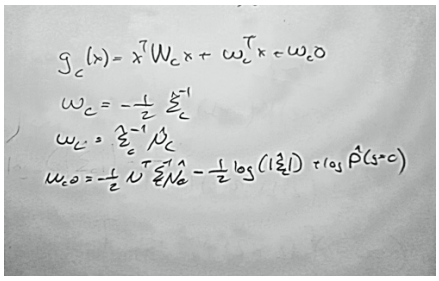
Parameter Estimations

After collecting the data, their parameters must be found. Sample means is a 2x3 array. First dimension is the mean of each feature of the selected class, second dim is the class. Each class has 2 mean values of the feature. “data[:,i]” is the data set of the i’th feature “np.mean(data[:,0][y==c+1])” is for finding the mean of the feature, c is the class label. Sample covariances “np.cov(data[:,0][y==c+1]” gives covariances of the X1 and combining X2 and X1 gives the covariances matrix. Class priors are the division of their sizes and total point sizes.



Score Values

For finding best estimation, the logarithmic posterior is used, and the constant are deleted since it’s useless for comparisons. The score function is as following with the W parameters. According to the score distribution, the class will be located for points.



Solution

A pandas packages figure called crosstab allows analyzing the predictions with the giving table. Each colon gives the distribution of the parametric methods solution. For example number (1,2) gives the number of the misprediction of the class 1 while claiming it as class 2. distribution values array has each classes score function. “C1[(C1 < C2) & (C1 < C3)] = np.nan” claims that C1 values are re-organized if the values of C1 are smaller than other class, the low values are deleted. “plt.contour(X1, X2, discriminant_values[:,:,0] - discriminant_values[:,:,1], levels = 0, colors = "k")” draws the line for class1 and class 2 separation where the probability of the class3 is lowest therefore its values are unnecessary

comparation.” C1[(C1 < C2) | (C1 < C3)] = np.nan” gives the only location where first class is highest.