WORK ORIGINATION CERTIFICATION

By submitting this document, I, **Dhruthi Sridhar Murthy**, the author of this deliverable, certify that

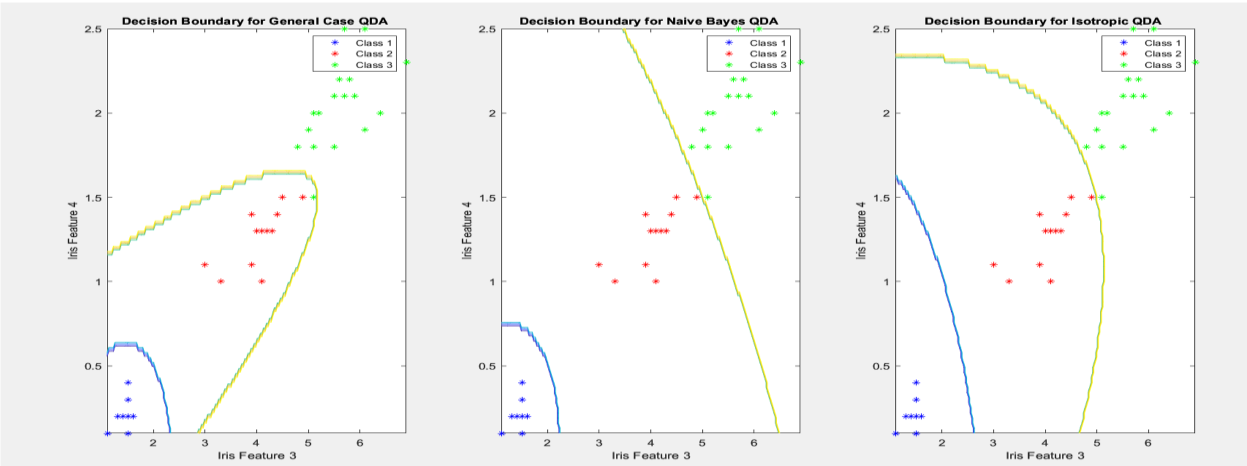
1. I have reviewed and understood Regulation UCF 5.015 of the current version of UCF's Golden Rule Student Handbook available at http://goldenrule.sdes.ucf.edu/docs/goldenrule.pdf, which discusses academic dishonesty (plagiarism, cheating, miscellaneous misconduct, etc.)
2. The content of this Major Project report reflects my personal work, and, in cases, it is not, the source(s) of the relevant material has/have been appropriately acknowledged after it has been first approved by the course's instructional staff.
3. In preparing and compiling all this report material, I have not collaborated with anyone, andI have not received any type of help from anyone but the course's instructional staff.

Signature **Dhruthi Sridhar Murthy** Date **10/27/2022**

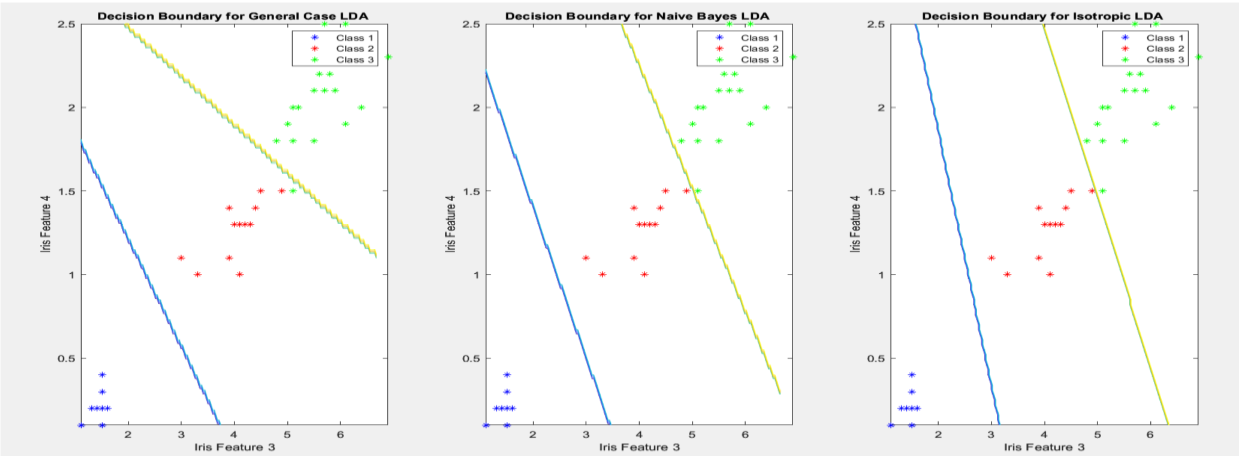
Task 2: (a),(b),(c)

There are 3 classes class 1, class 2, and class 3 namely Sentosa, Versicolor, and Virginia. The graph below represents the decision boundaries for QDA in the general case, Naïve Bayes, and isotopic analysis.

When the gaussian matrix took various covariance matrices for different models we get the below plots for QDA.



When the Gaussian matrix has the same shared covariance matrix for different models, we get the below plot for linear discrete analysis.



Model	Classifier	Test Set Error
General case	LDA	0.06
Naïve Bayes	LDA	0.05
Isotropic	LDA	0.04
General case	QDA	0.04
Naïve case	QDA	0.04
Isotropic	QDA	0.04

The about table displays the different test set errors for different models with two classifiers (QDA and LDA).

Question 1 :

Task 1 :

(a) Training set $\{(x_n, l_n)\}_{n=1}^{N}$

c-class classification problem — N_K samples from class k. $K = 1, 2, \ldots, C$, Maximum likelihood estimator (MLEtor) for the covariance matrix C.

Class-conditional densities $p(x|N_K)$

Prior class probabilities $p(N_K)$

Consider first the case of two classes, each having a gaussian class-conditional density with a shared covariance matrix, we have a data set $\{x_n, t_n\}$ where, $n = 1, \ldots N$

Training set & variation set is got split into data from data set $\{x_n, t_n\}$.

Here, $t_n = 1$ denotes class N_1 and $l_n = 0$ denotes class N_2. We denote the prior class probability $p(N_1) = \pi$, so that $p(N_2) = 1 - \pi$. for a data point x_n from class N_1, we have $l_n = 1$ & hence,

$$p(x_n, N_1) = p(N_1)p(x_n|N_1) = \pi N(x_n|\mu_1, \Sigma)$$

Similarly for class N_2, we have $l_n = 0$ and hence

$$p(x_n, N_2) = p(N_2)\,p(x_n|N_2)$$
$$= (1-\pi)\,N(x_n|\mu_2, \Sigma)$$

$$p(t, x|\pi, \mu_1, \mu_2, \Sigma) = \prod_{n=1}^{N} [\pi N(x_n|\mu_1, \Sigma)]^{t_n} [(1-\pi)N(x_n|\mu_2, \Sigma)]^{1-t_n}$$

where, $t = (t_1, \ldots t_N)^T$. consider first the maximization with respect to π. The terms in the log likelihood function that depend on π are,

$$\sum_{n=1}^{N} \{ t_n \ln \pi + (1-t_n) \ln(1-\pi) \}$$

setting the derivative with respect to π equal to zero & rearraging, we obtain.

$$\pi = \frac{1}{N} \sum_{n=1}^{N} t_n = \frac{N_1}{N} = \frac{N_1}{N_1 + N_2}$$

where, N_1 denotes the total number of data points in class N_1 and N_2 denotes the total number of data points in class N_2.

Thus the maximum likelihood estimate for π is simply the fraction of points in class N_1 as expected. This result is easily generalized to the multiclass case where again the maxium likelihood estimate of the prior probability associated with class N_K is given by the fraction of the training set points assigned to that class.

Now consider the maximization with respect to μ_1. Again we can pick out of the log likelihood function those terms that depend on μ_1 giving.

$$\sum_{n=1}^{N} t_n \ln N(x_n | \mu_1, \Sigma) = -\frac{1}{2} \sum_{n=1}^{N} t_n (x_n - \mu_1)^T \Sigma^{-1} (x_n - \mu_1) + \text{const.}$$

setting the derivative with respect to μ_1 to zero & rearranging, we obtain.

$$\mu_1 = \frac{1}{N_1} \sum_{n=1}^{N} t_n x_n$$

which is simply the mean of all the input vectors x_n assigned to class C_1. By a similar argument, the corresponding result for μ_2 is given by

$$\hat{\mu}_2 = \frac{1}{N_2} \sum_{n=1}^{N} (1-t_n) x_n$$

which again is the mean of all the input vectors x_n assigned to class C_2.

Finally, consider the maximum likelihood solution for the shared covariance matrix Σ. Picking out of the terms in the log likelihood function that depend on Σ, we have

$$-\frac{1}{2} \sum_{n=1}^{N} t_n \ln|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} t_n (x_n - \mu_1)^T \Sigma^{-1} (x_n - \mu_1)$$

$$-\frac{1}{2} \sum_{n=1}^{N} (1-t_n) \ln|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (1-t_n)(x_n - \mu_2)^T \Sigma^{-1} (x_n - \mu_2)$$

$$= -\frac{N}{2} \ln|\Sigma| - \frac{N}{2} \mathrm{Tr}\{\Sigma^{-1} S\}$$

where, we have defined.

$$S = \frac{N_1}{N} S_1 + \frac{N_2}{N} S_2$$

$$S_1 = \frac{1}{N_1} \sum_{n \in C_1} (x_n - \mu_1)(x_n - \mu_2)^T$$

$$S_2 = \frac{1}{N_2} \sum_{n \in C_2} (x_n - \mu_2)(x_n - \mu_2)^T$$

Using the standard result for the maximum likelihood solution for a gaussian distribution, we see that $\Sigma = S$, which represents a weighted average of the covariance matrices associated with each of the two classes separately.

This result is easily extended to the K class problem to obtain the corresponding maximum likelihood solutions for the parameter in which each class conditional density is gaussian with a shared covariance matrix. Note that the approach of fitting gaussian distributions to the classes is not robust to outliers, because the maximum likelihood estimation of a gaussian is not robust.

(b) MVG data : (Mean)

$$nl(\mu, C|D) = \frac{N}{2}\left[D\ln(2\pi) + \ln\det(C) + \text{trace}\{C^{-1}\hat{C}\} + \|\mu - \hat{\mu}\|^2_{C^{-1}}\right] \quad — ①$$

According to MLE, estimating the MVG density amounts to solving the constrained minimization problem

$$\inf_{\substack{\mu \in \mathbb{R}^D \\ C \in \Omega_C}} nl(\mu, C|D)$$

~~The equation (1) can be estimating the MVG density amounts to solving the constrained minimization problem,~~

The negative log-likelihood in (1) can be shown to be jointly convex in both μ and C.

MLEtor of the Mean Vector :

- If $\hat{C}_{MLE} > 0$, then $\|-\|_{\hat{C}^{-1}_{MLE}}$ is indeed a weighted Euclidean norm & , $\therefore$ we see from (30) that,

$$\boxed{\hat{\mu}_{MLE} \equiv \hat{\mu}} \quad — ②$$

- $\exists\, \hat{C}^{-1}_{MLE}$, then $\hat{\mu}_{MLE}$ is unique & coincides with the sample mean of the training set D.

eqn ② is given by MLEtor.

· MLEtor of the covariance Matrix : uncorrelated Variates/Independent

Assuming that the data in D came from an MVG with uncorrelated or independent variates, then $C = diag(v)$ for some $v \geq 0$, or equivalently stated in terms of the precision matrix, $A = diag(a)$ for some $a \geq 0$. Hence, .

$$a \in \Omega_a \triangleq \{a \in \mathbb{R}^D : a \geq 0\}$$

$\mu = \hat{\mu}_{MLE} = \hat{\mu}$ from eq^n ① becomes,

$$\inf_{a \in \Omega_a} l(\hat{\mu}, [diag(a)]^{-1} | D) \iff \inf_{a \in \Omega_0} \frac{N}{2}\left[D \ln(2\pi) - \sum_{d=1}^{D} \ln a_d + a^T diag(\hat{c})\right] \quad \text{— ③}$$

For ③ the stationary point equation for a becomes,

$$\left.\frac{dl(\hat{\mu}, [diag(a)]^{-1} | D)}{da}\right|_{a = \hat{a}_{MLE}} = 0 \iff \hat{a}_{MLE} = [diag(\hat{c})]^{-1}$$

$$\iff \hat{C}_{MLE} = diag(diag(\hat{c})) \quad \text{— ④}$$

(c) MLEtor of the covariance Matrix : Isotropic

If we assume that the data in D came from an isotropic MVG, then $C = vI$ for some $v \geq 0$ or equivalently stated in terms of the precision matrix,

$$A = \frac{1}{a} I \qquad a \triangleq \frac{1}{v} \geq 0$$

New parameter,

$$a \in \Omega_a \triangleq \{a \in \mathbb{R} : a \geq 0\}$$

set, $\mu = \hat{\mu}_{MLE} = \hat{\mu}$ from eqn① becomes,

$$\inf_{a \in \Omega_a} l\left(\hat{\mu}, \frac{1}{a} I \mid D\right) \Leftrightarrow \inf_{a \in \Omega_a} \frac{N}{2}\left[D \ln(2\pi) - D \ln a + a \operatorname{trace}\{\hat{C}\}\right] \quad — (5)$$

* MLEtor of the covariance Matrix: Isotropic

data D came from an isotropic MVG, then $C = VI$, $V \geq 0$

precision matrix, $\Lambda = \frac{1}{a} I$, $a \triangleq \frac{1}{V} \geq 0$.

$$a \in \Omega_a \triangleq \{a \in \mathbb{R} : a \geq 0\}$$

$\mu = \hat{\mu}_{MLE} = \hat{\mu}$ — eq^n (2) becomes,

$$\inf_{a \in \Omega_a} l\left(\hat{\mu}, \frac{1}{a} I \mid D\right) \Leftrightarrow \inf_{a \in \Omega_a} \frac{N}{2}\left[D \ln(2\pi) - D \ln a + a \operatorname{trace}\{\hat{C}\}\right] \quad — (6)$$

From (6) the stationary point equation for a becomes,

$$\left.\frac{dl\left(\hat{\mu}, \frac{1}{a} I \mid D\right)}{da}\right|_{a = \hat{a}_{MLE}} = 0 \Leftrightarrow \hat{a}_{MLE} = \left[\frac{1}{D} \operatorname{trace}\{\hat{C}\}\right]^{-1}$$

$$\Leftrightarrow \hat{C}_{MLE} = \left(\frac{1}{D} \operatorname{trace}\{\hat{C}\}\right) I \quad — (7)$$