ECE5258 Patter Recognition (Fall 2020)

MP1: LDA, QDA & RQDA classifiers

Mawaba Pascal Dao

Dr. Anagnostopoulos

Oct 14, 2020

Git Repo: https://github.com/PascalPolygon/pattern_recognition/blob/master/MP1/MP1.ipynb

1. Task 1

a/ Assuming that we have already estimated the class specific $\hat{C}_k$ in the QDA scenario, we can compute C_{pooled} as a weighted sum of the $\hat{C}_k$; $k \in \{1, 2, \dots c\}$

This is because for LDA, $\hat{C}_{MLE} = C_k$.

Let $\hat{P}_k$; $k \in \{1, 2, \dots c\}$ be the class priors for each class. Then we express the weighted sum of class specific $\hat{C}_k$'s as:

$$\boxed{\hat{C}_{pooled} = \sum_{k=1}^{c} \hat{P}_k \hat{C}_k}$$

b/ If we assume the data in D come from an MVG with uncorrelated independent variates, the $C = diag(v)$ for some $v \geq 0$, equivalently $C^{-1} = A = diag(a)$ for some $a \geq 0$ and the negative log-likelihood becomes

$$Nl(\hat{M}, [diag(a)]^{-1} | D)$$

a/ Assuming that we have already estimated the class specific $\hat{C}_k$ in the QDA scenario, we can compute C_{pooled} as a weighted sum of the $\hat{C}_k$; $k \in \{1, 2, \dots c\}$

This is because for LDA, $\hat{C}_{MLE} = C_k$.

Let $\hat{P}_k$; $k \in \{1, 2, \dots c\}$ be the class priors for each class. Then we express the weighted sum of class specific $\hat{C}_k$'s as:

$$\boxed{\hat{C}_{pooled} = \sum_{k=1}^{c} \hat{P}_k \hat{C}_k}$$

b/ If we assume the data in D come from an MVG with uncorrelated independent variates, the $C = diag(v)$ for some $v \geq 0$, equivalently $C^{-1} = A = diag(a)$ for some $a \geq 0$ and the negative log-likelihood becomes

$$Nl(\hat{M}, [diag(a)]^{-1} | D)$$

2. Task 2

a. QDA

i. Training samples

ii. Means of training samples

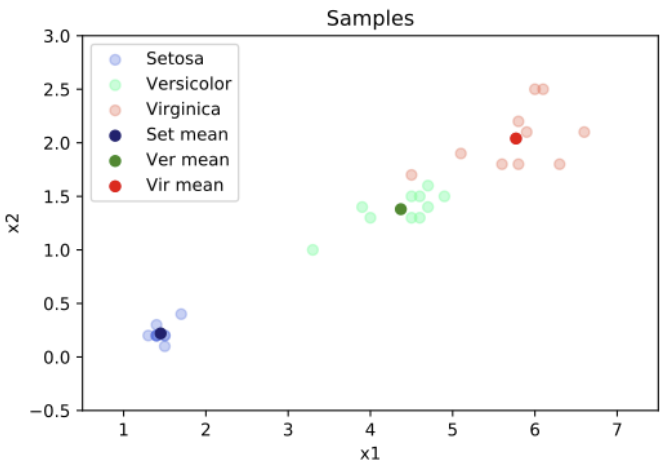

Figure 1 Color coded training samples and their means

iii. Tolerance Region Plots (50%-content, 99%-coverage)

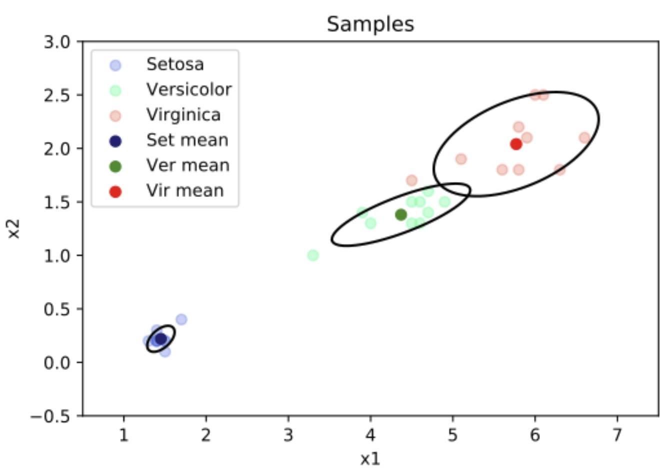

Figure 2 Tolerance regions for QDA with general samples

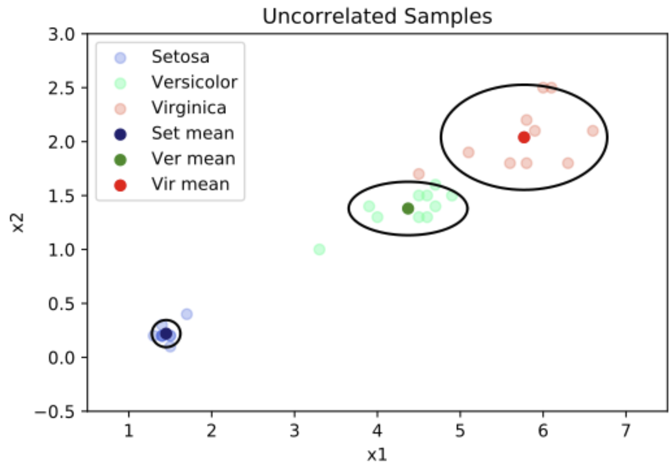

Figure 3 Tolerance regions for QDA with uncorrelated samples

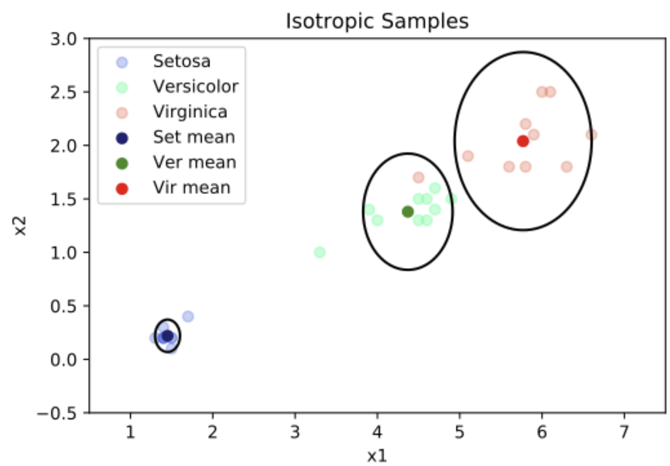

Figure 4 Tolerance regions for QDA with isotropic samples

iv. Decision Region Plots

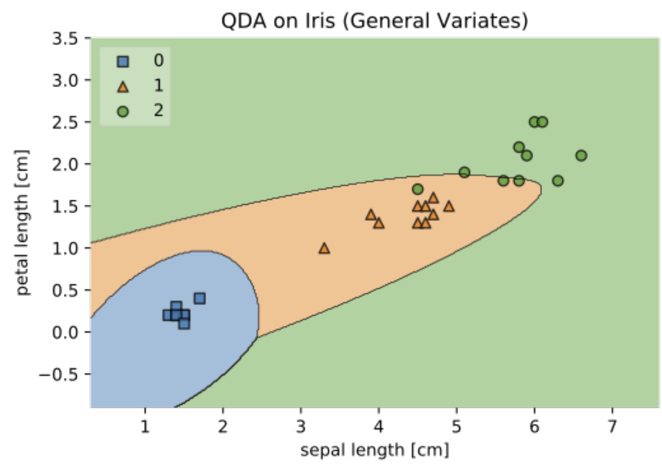

Figure 5 Decision regions for QDA with general samples

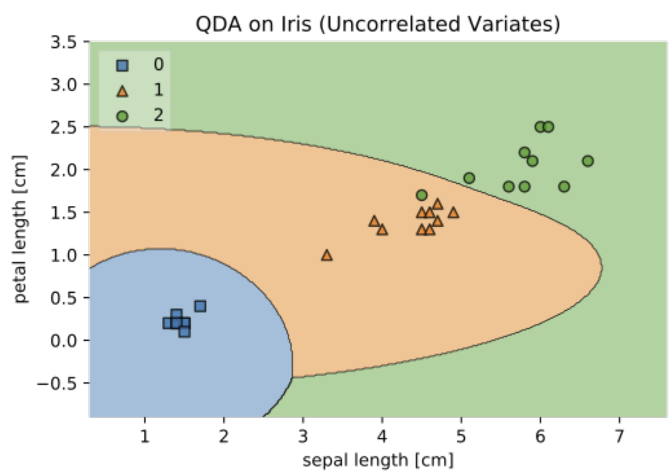

Figure 6 Decision regions for QDA with uncorrelated samples

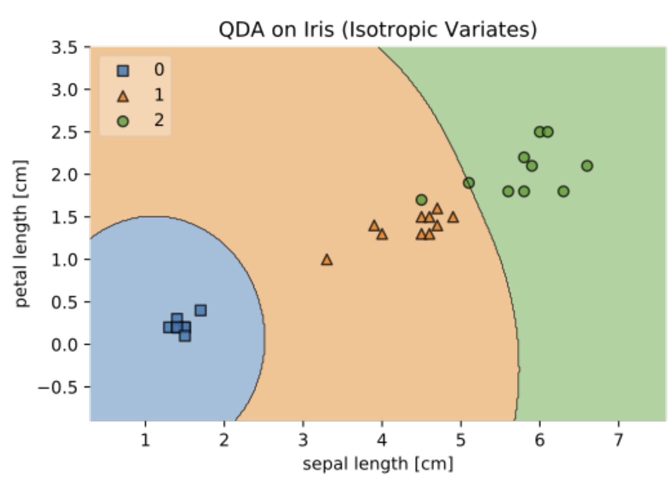

Figure 7 Decision regions for QDA with isotropic samples

Overall the classifier does a good job of discriminating between samples from the training data. From looking at the previous tolerance region plot, we notice the overlap between the region of class 1 (Versicolor) and that of class 2 (Virginica). It was therefore expected that the model might misclassify some Virginica samples (those close to the decision boundary) as coming from the Versicolor batch.

b. LDA

LDA has the same samples and means as QDA, therefore the color-coded plot depicting the training samples and means is skipped

i. Tolerance Region plots

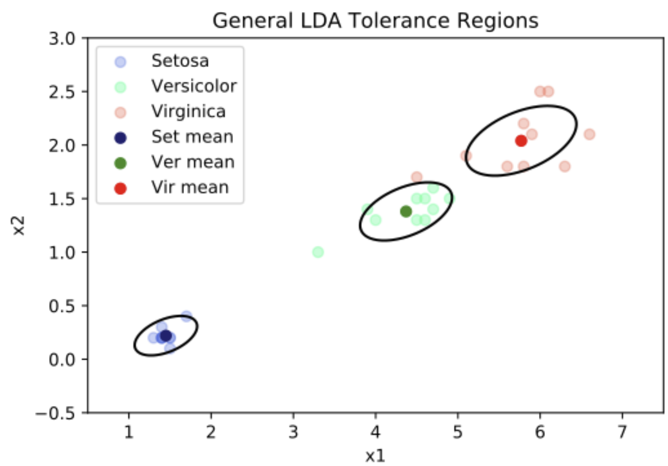

Figure 8 Tolerance regions for LDA with general samples

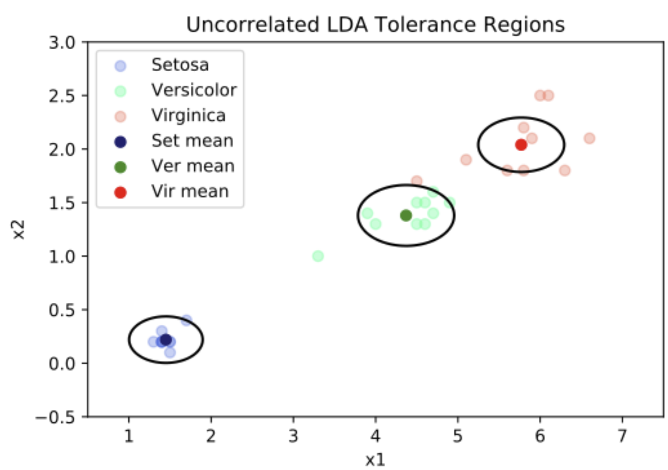

Figure 9 Tolerance regions for LDA with uncorrelated samples

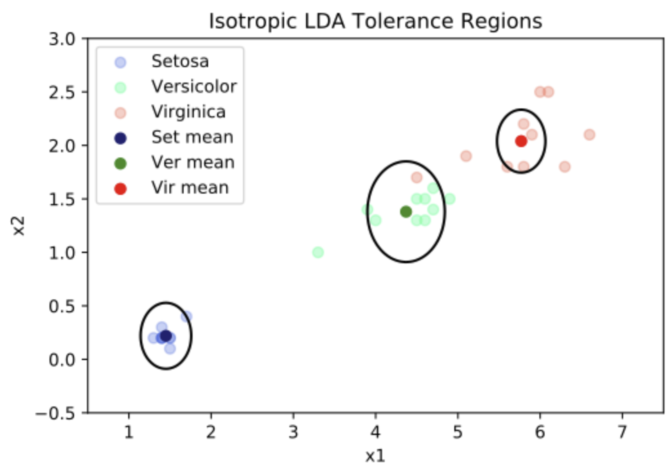

Figure 10 Tolerance regions for LDA with isotropic samples

ii. Decision Region Plots

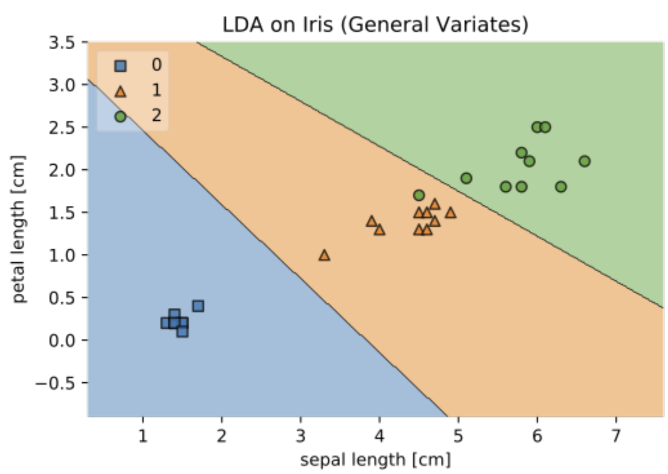

Figure 11 Decision regions for LDA with general samples

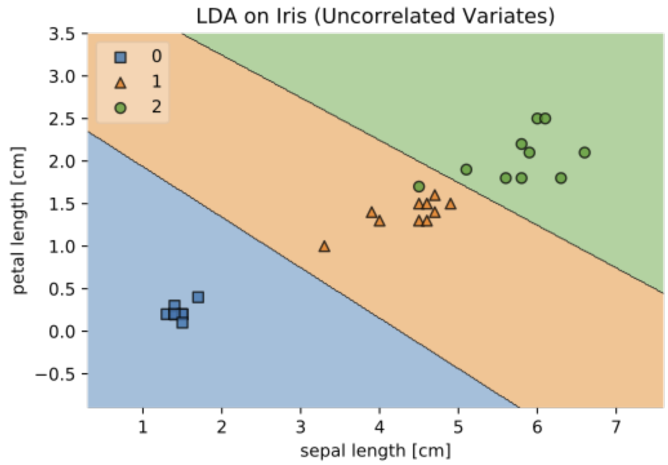

Figure 12 Decision regions for LDA with uncorrelated samples

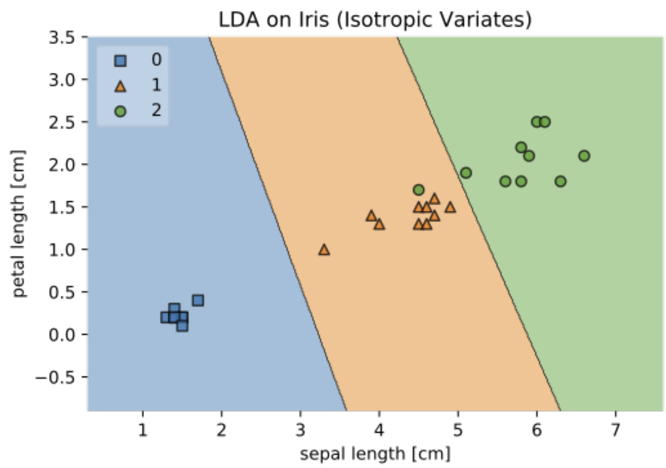

Figure 13 Decision regions for LDA with isotropic samples

The LDA classifier performs better on the training set than did QDA. Both classifiers do a good job at discriminating between samples, but for the most ambiguous cases (The Virginica samples at the decision boundary with the Versicolor samples), the LDA model only misclassifies one Virginica training sample as coming from the Versicolor regardless of the covariance matrix used, while with QDA, in the general case, there were 3 misclassification errors.

c. Finding best classifier
    i. MR on training set

General QDA	General LDA	Uncorrelated QDA	Uncorrelated LDA	Isotropic QDA	Isotropic LDA
0.1	0.033	0.033	0.033	0.033	0.033

Table 1 Misclassification rate on training set

ii. MR on validation set

General QDA	**General LDA**	Uncorrelated QDA	**Uncorrelated LDA**	Isotropic QDA	Isotropic LDA
0.1	**0.033**	0.067	**0.033**	0.1	0.0833

Table 2 Misclassification rate on validation set

The best classifier per its validation set performance is the LDA classifier with general or uncorrelated covariance matrix. When using this model to predict labels on the test set, we get a misclassification rate of 0.05, which is twice as good as QDA on the training set. This result is consistent with the tolerance and decision regions computed. The tolerance regions for the LDA classifier were smaller than those on the QDA classifier for the same coverage and content values as well as for any covariance matrix case. This means that the LDA tolerance regions were better centered around the mean of the distribution and less likely to include samples from a different distribution. Furthermore, when using general or uncorrelated covariance, the shape of the tolerance region is such that the classifier avoids the one outlier sample (in training set) from class 2 (Virginica), which is labeled as class 1 (Versicolor) by the other classifiers. LDA with general or uncorrelated covariance matrix outperforms the other classifiers because it is more conservative in its tolerance and decision regions/rules, which is an advantage on the data at hand. The Honest estimate of the best model on the test set is tabulated below.

General LDA	Uncorrelated
0.05	0.05

Table 3 Honest estimate of best model on test set

3. Task 3

a. RQDA plots

In the following plots:

0 = Setosa class

1 = Versicolor class

2 = Virginica class

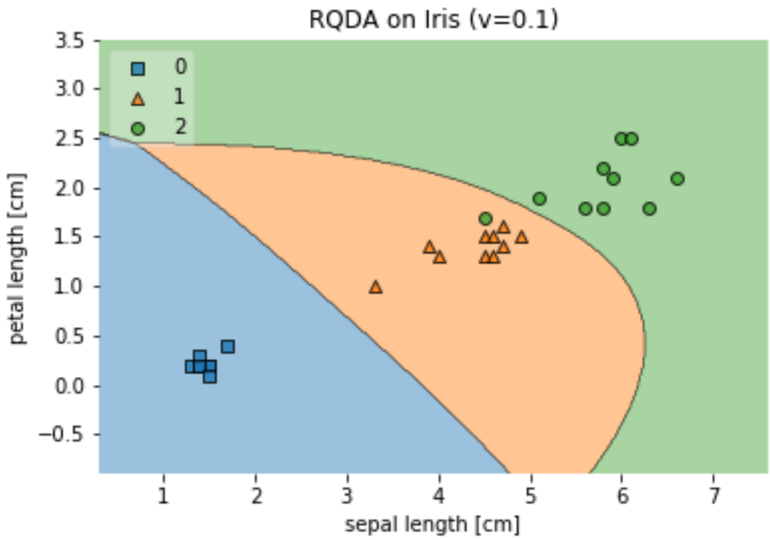

Figure 14 RQDA decision regions plot

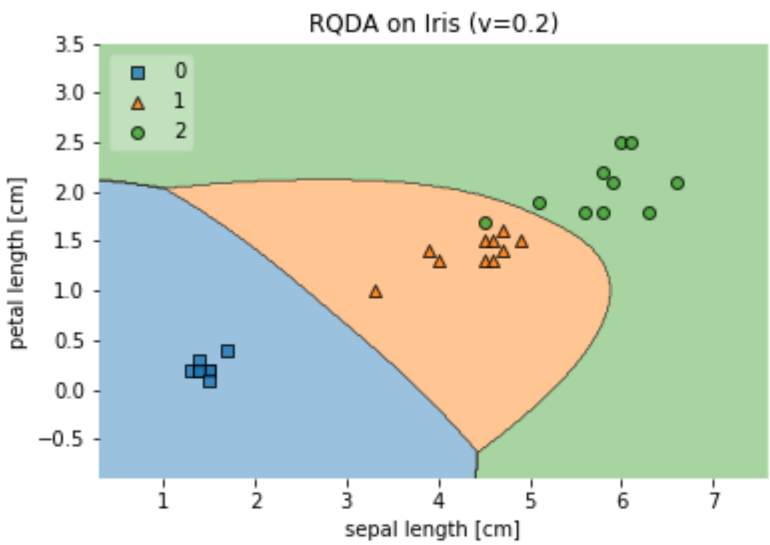

Figure 15 RQDA decision regions plot

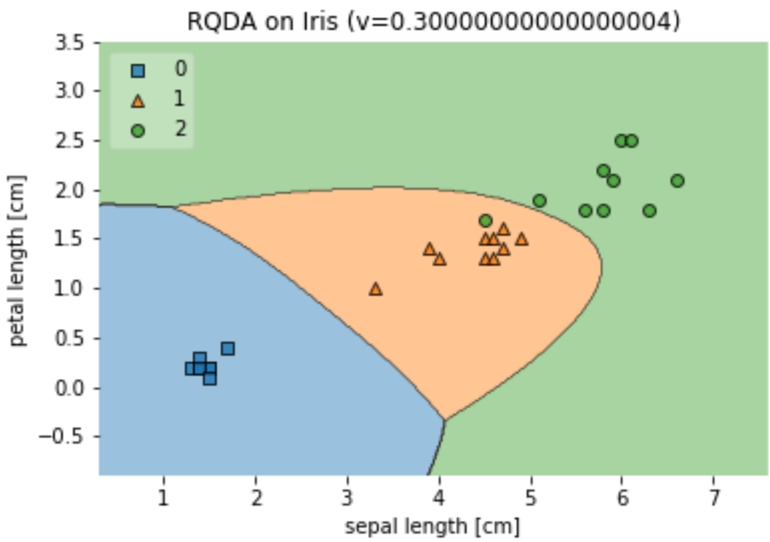

Figure 16 RQDA decision regions plot

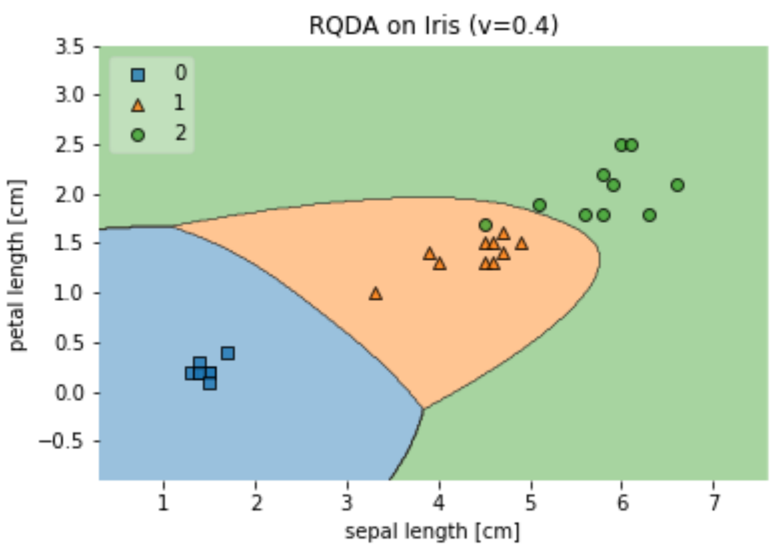

Figure 17 RQDA decision regions plot

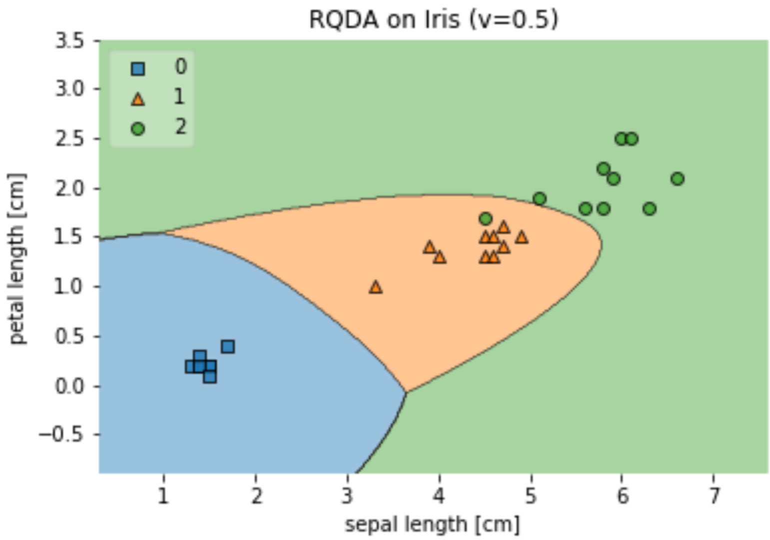

Figure 18 RQDA decision regions plot

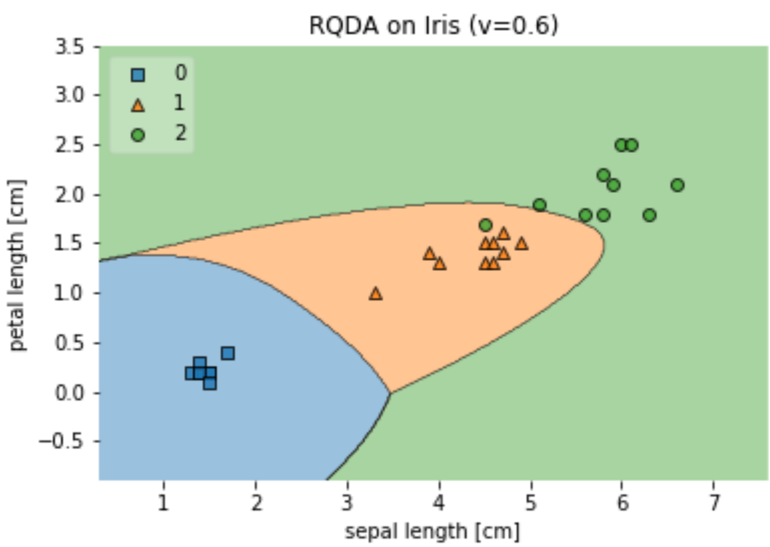

Figure 19 RQDA decision regions plot

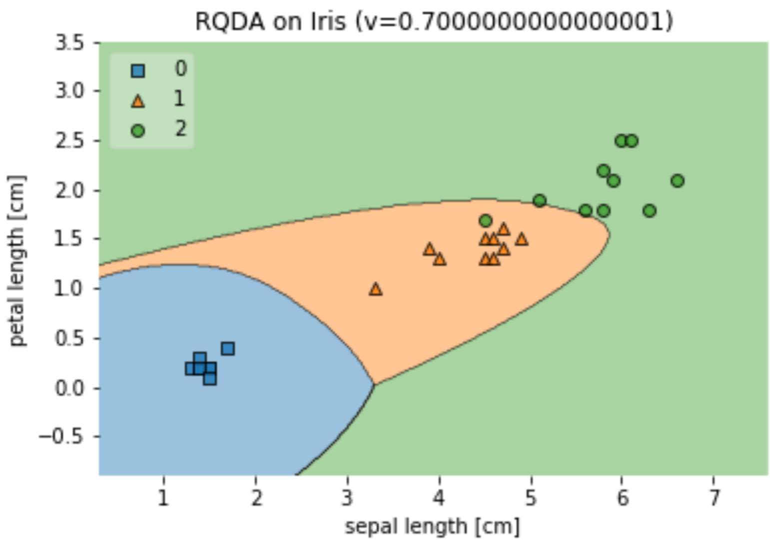

Figure 20 RQDA decision regions plot

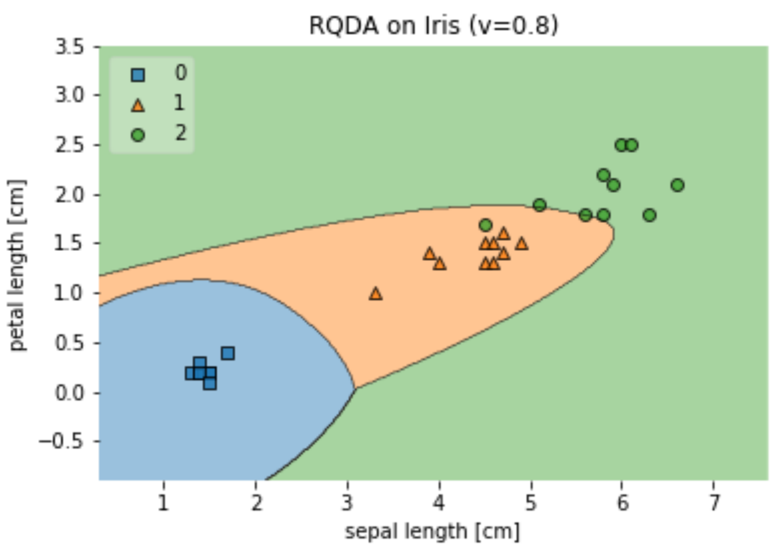

Figure 21 RQDA decision regions plot

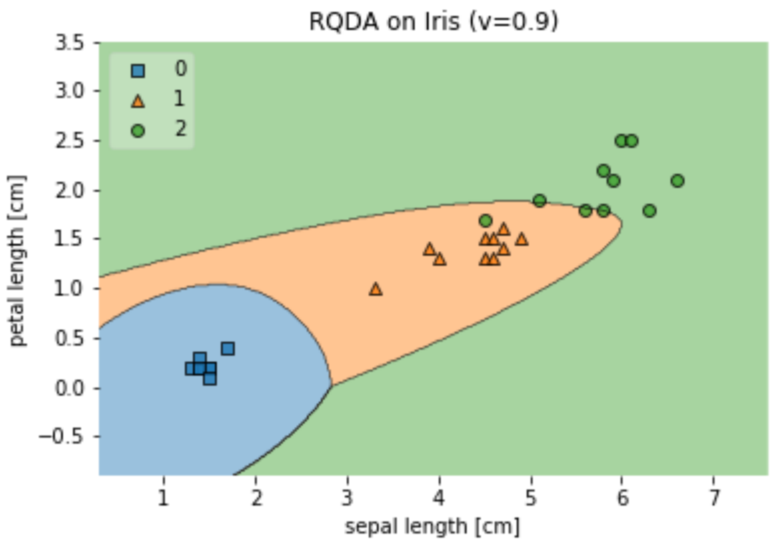

Figure 22 RQDA decision regions plot

$$C_{RQDA} = \breve{C}_k = v\hat{C}_k + (1-v)\hat{C}_{pooled}$$

The expression for computing the RQDA covariance matrix is a linear interpolation between the QDA and LDA covariance matrices. When v=0, C_{RQDA}=C_{pooled} and the decision region becomes that of an LDA classifier and when v=1, C_{RQDA}=C_k and the decision region is that of a QDA classifier. The plots confirm this, as we slide v from 0.1 to 0.9 the decision region plots which starts looking closer to the one for LDA previously seen morphs into a decision region closer to QDA for higher values of v. Where *v=0.5* would be the midpoint between an LDA and QDA decision regions.

b. RQDA vs QDA/LDA

Finding best RDQA classifier

v	Misclassification Rate
0.1	0.067
0.2	0.067
0.3	0.067
0.4	0.067
0.5	0.067
0.6	0.067
0.7	0.083
0.8	0.083
0.9	0.1

Table 4 RQDA performance for different v values

Unsurprisingly, for smaller values of v the misclassification error is smaller. When v is close to 0 the RQDA has more LDA properties, the misclassification error is similar to that obtained with LDA. As seen in previous results, LDA outperforms QDA on the iris data, therefore RQDA that mimics LDA closely has better results than RQDA that mimics QDA. The best RQDA classifier is when $v=0.1$. **When using this RQDA classifier with v=0.1 on the test data we get the same misclassification rate as with LDA, 0.05.**