STOR 565 Fall 2019 Homework 6

Due on 01/31/2018 in Class

Hunter Finger

Remark. Credits for **Theoretical Part** and **Computational Part** are in total 100 pt. For **Computational Part**, please complete your answer in the **RMarkdown** file and summit your printed PDF homework created by it.

##Comment If dplyr and MASS are both loaded, you might need to specify `dplyr::select` to specify that you want the dplyr version of the `select` function.

Computational Part

###About the data: Tree leaf images

We will attempt to identify trees based on image data of their leaves. This is a tough problem, though apps such as iNaturalist now do a pretty good job identifying plants from images taken on your phone.

The data set is from here: https://www.kaggle.com/c/leaf-classification/data

Images have been pre-processed, so the dataset inlcudes vectors for margin, shape and texture attributes for each of almost 1000 images. We will focus on the shape attributes, which describe the contours of the leaf in the image.

###A helpful demonstration for SVM

http://uc-r.github.io/svm

###Q1 ###(a) (3 points)

Load the `leaf_train` dataset.

(i) Subset the columns to include only `id`, `species` and the `shape` variables, which is most easily done using the dplyr `select` function and the sub-function `contains`. There should be 66 variables in all.

```
leaf_train = read.csv("leaf_train.csv", stringsAsFactors = F)
leaf = select(leaf_train, id, species, contains("shape"))
```

(ii) Then create a new variable `genus` by extracting the first part of the species name. You can use the following code, assuming your data objects are named in a compatible way. You will probably want to load the data with `stringsAsFactors` as false.

(iii) Lastly, convert the genus variable to a factor.

```
leaf$genus <- str_split(leaf$species, "_", simplify = TRUE)[, 1]
leaf$genus = as.factor(leaf$genus)
```

(iv) Display your resulting data frame and the result of `summary(leaf$genus)`, which should give the number of observations of each genus. **Display only the id, species and first two species variables in your output, and only five rows of the data, eg by using the head function.**

```
head(leaf[,1:4], n = 5)
```

```
##   id                species     shape1     shape2
## 1  1            Acer_Opalus 0.00064671 0.00060945
## 2  2 Pterocarya_Stenoptera 0.00074942 0.00069461
## 3  3  Quercus_Hartwissiana 0.00097311 0.00091025
## 4  5       Tilia_Tomentosa 0.00045312 0.00046534
## 5  6    Quercus_Variabilis 0.00068161 0.00059775
```

```
summary(leaf$genus)
```

```
##          Acer        Alnus  Arundinaria       Betula   Callicarpa
##           100           50           10           20           10
##      Castanea       Celtis       Cercis       Cornus      Cotinus
##            10           10           10           30           10
##     Crataegus      Cytisus   Eucalyptus        Fagus       Ginkgo
##            10           10           30           10           10
##          Ilex  Liquidambar Liriodendron  Lithocarpus     Magnolia
##            20           10           10           20           20
##         Morus         Olea  Phildelphus      Populus       Prunus
##            10           10           10           30           20
##    Pterocarya      Quercus Rhododendron        Salix       Sorbus
##            10          380           10           20           10
##        Tilia        Ulmus     Viburnum      Zelkova
##            30           10           20           10
```

(v) Randomly split your data into test and training sets. About 35 percent of the data should be in the test set. Display a summary of genus labels in the training set.

Note: In the rare event that one class in the training data is not represented, you may reduce the test set percentage to 30 percent and resample.

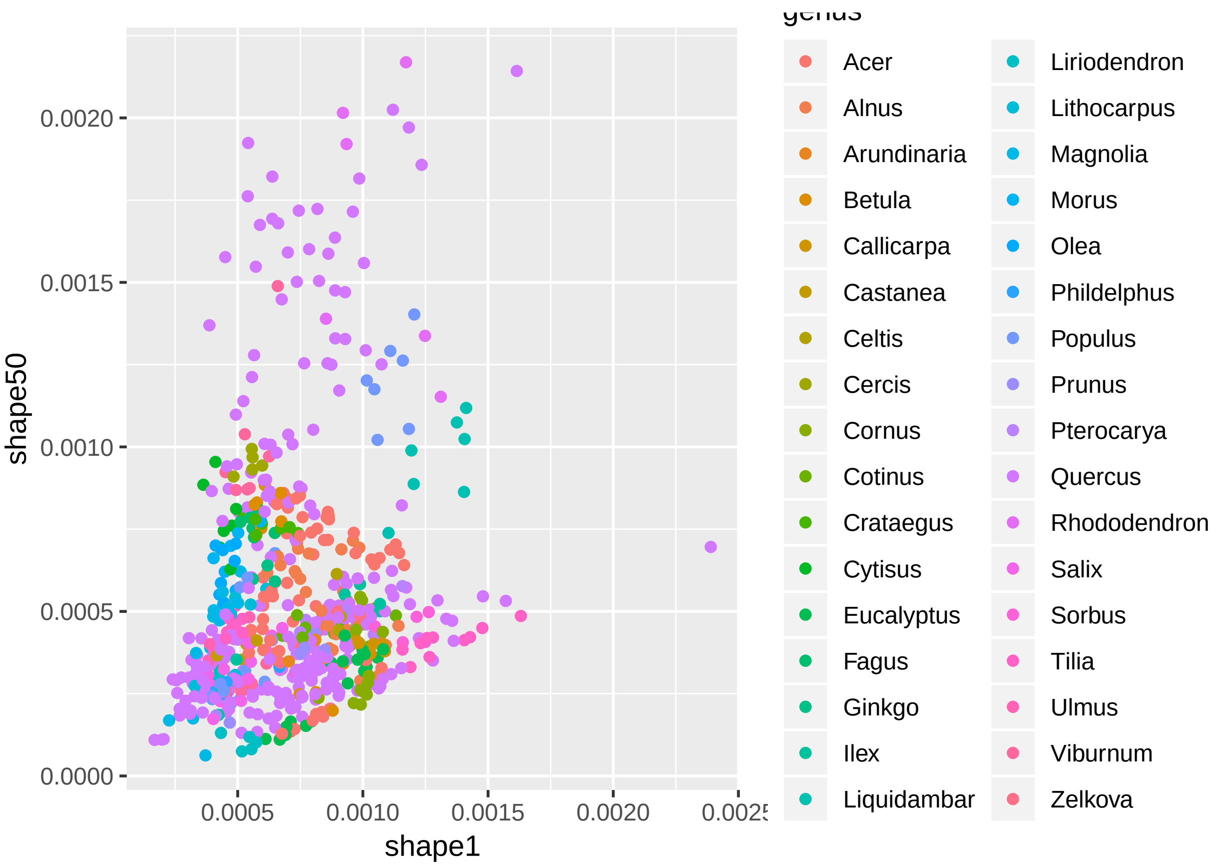
```
train = sample(1:dim(leaf)[1], dim(leaf)[1]/100*65)
test = -train
leaf_train = leaf[train,]
leaf_test = leaf[test,]
```

##(b) (2 points)

For the training data:

(i) Make a scatter plot of shape1 by shape50, with some form of genus label. ggplot2 is probably the best package for this, though you do not need to make the plot fancier than required to display the information above.

```
ggplot(leaf_train, aes(x = shape1, y = shape50, color = genus)) + geom_point()
```



(ii) Write two to three sentences discussing some possible implications of this plot for the SVM model.

ANSWER Having 34 classes in which the model has to differentiate might be difficult. In this example it would be fairly simple to create a hyperplane that seperated two classes, but we have 34 classes with many shapes. The model with do what we can simply see here, but for all features and all classes.

##(c) (15 points)

For the training data:

(i) Write a function, or use an available one, to choose the cost parameter for the SVM model on this training data with **linear kernel.** Use **shape variables as predictors only, genus as response**.

Use **5-fold cross validation.** Use the array of costs provided in the code below.

If you use a built-in function, you must state specifically how the best parameter value is chosen, for example by giving the error function minimized. Simply stating `classification error` is insufficient and will receive no points. You must state what that means. If using your own function, you may use any error function you like that is justified for classification problems.

See the demo linked above for help.

This might take some time to run. Do not knit your file at the last minute before the assignment is due.

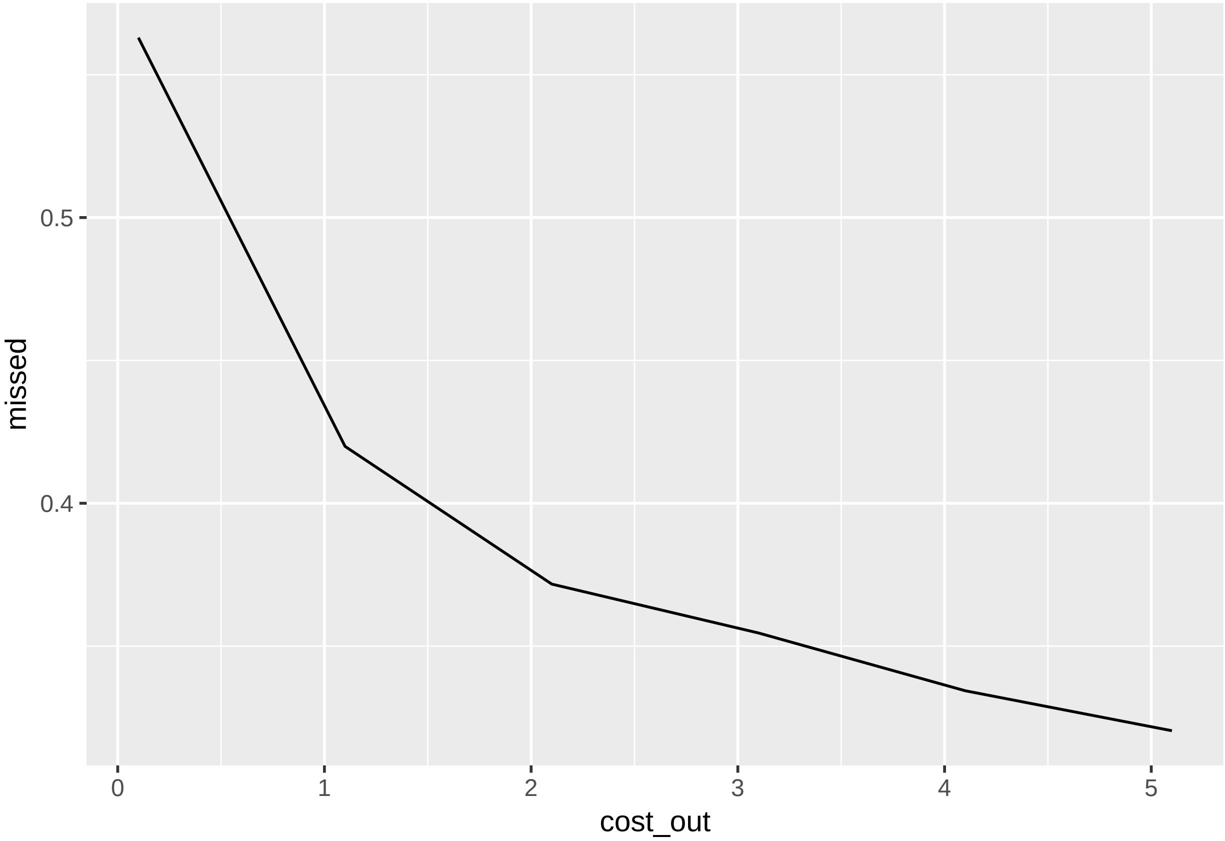
(ii) Report the best value of cost chosen, and plot the errors by the cost values.

(iii) Write two or three sentences discussing some basic implications of your answer in (ii), using the concepts from class. Lecture 7 will be helpful.

```
cost_out <- seq(from = 0.1, to =5.1, by = 1)
missed = rep(NA, length(cost_out))
```

```
for(i in 1:length(cost_out)){
  svm.mod = svm(formula = genus~., data = select(leaf_train, -c(id, species)), kernel = "linear", cost =
  missed[i] = (sum(svm.mod$fitted != leaf_train$genus))/nrow(leaf_train)
}

errors = data.frame(cbind(cost_out, missed))
```

```
ggplot(errors, aes(x = cost_out, y = missed)) + geom_line()
```



ANSWER 5.1 is the best cost value for the model. It has the lowest MSE of the 5 options. The value of c is accounting for how much noise or values are on the wrong side of the plane. As we saw in our earlier plot it is non trivial to accomplish this so having a higher c will help us make more accurate predictions on our messy data.

##(d) (15 points)

(i) Run the SVM model on the **training data** with **linear kernel** and the cost determined in part (c). If you are unable to do part (c), use a cost of 1, the default. Report a summary of the fitted class label counts.

(ii) Create a classification plot from the model, plotting the variables `shape50` by `shape1`. See `?plot.svm`. In your plot statement, use the argument `xlim = c(0, 0.0012), ylim = c(0, 0.0012)`.

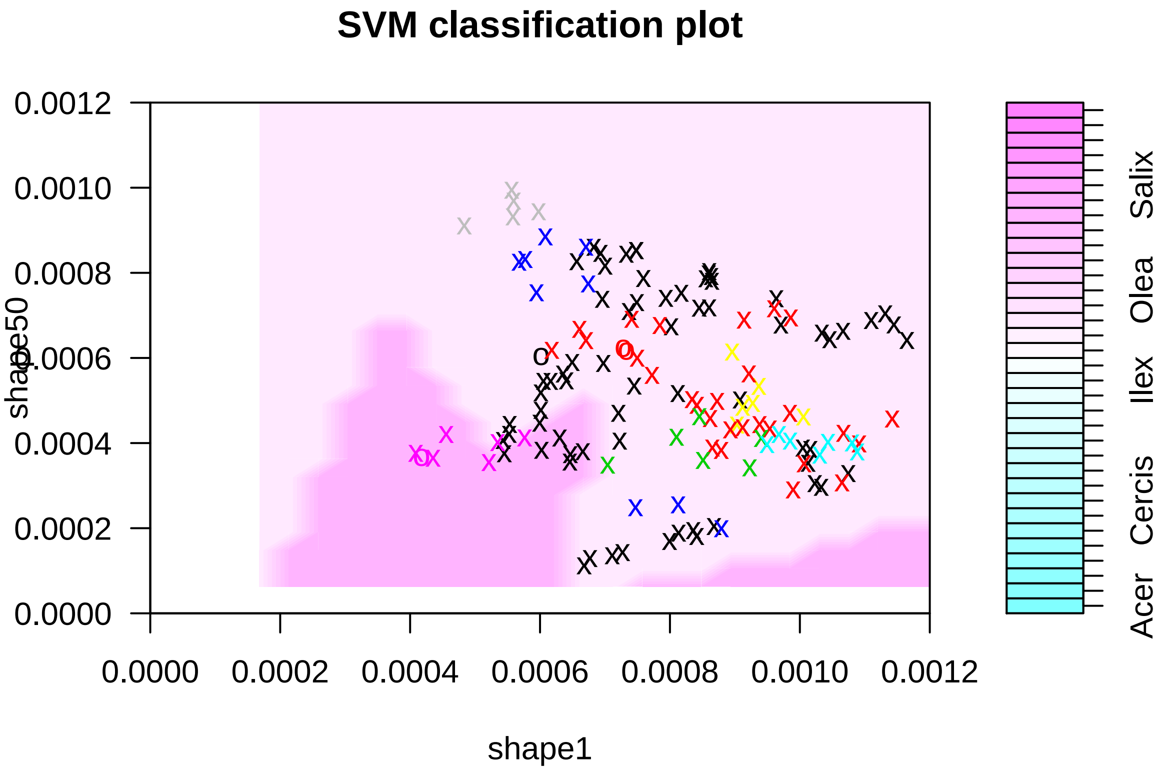
See the linked demo for an explanation of the plot. Write two sentences explaining what you see **using concepts and terminology from class.**

(iii) Predict outcomes based on your model in (i) for the test data. Display a confusion matrix and compute sensitivity, specificity statistics. You may use the function demonstrated in class.

Warning: the confusion matrix will be awkward to display. Don't worry about it so much.

The sensitivity and specificity are good summaries.

```
svm.mod = svm(formula = genus~., data = select(leaf_train, -c(id, species)), kernel = "linear", cost =5

plot(svm.mod, data = select(leaf_train, -c(id, species)), shape50~shape1, xlim = c(0, 0.0012), ylim = c
```



```
svm.pred = predict(svm.mod, newdata = leaf_test)
library(caret)
```

```
## Loading required package: lattice
```

```
confusion = confusionMatrix(leaf_test$genus, svm.pred)
confusion
```

```
## Confusion Matrix and Statistics
##
##                Reference
## Prediction     Acer Alnus Arundinaria Betula Callicarpa Castanea Celtis
##   Acer           14     0           0      0          0        0      0
##   Alnus           0     9           0      0          0        0      0
##   Arundinaria     0     0           0      0          0        0      0
##   Betula          0     0           0      0          0        0      0
##   Callicarpa      0     0           0      0          0        0      0
##   Castanea        0     0           0      0          0        1      0
##   Celtis          0     0           0      0          0        0      0
##   Cercis          0     0           0      0          0        0      0
##   Cornus          0     0           0      0          0        0      0
##   Cotinus         0     0           0      0          0        0      0
##   Crataegus       0     0           0      0          0        0      0
##   Cytisus         0     0           0      0          0        0      0
##   Eucalyptus      1     0           0      0          0        0      0
##   Fagus           0     0           0      0          0        0      0
##   Ginkgo          1     0           0      0          0        0      0
##   Ilex            0     0           0      0          0        0      0
##   Liquidambar     0     0           0      0          0        0      0
##   Liriodendron    0     0           0      0          0        0      0
##   Lithocarpus     0     0           0      0          0        0      0
##   Magnolia        0     0           0      0          0        0      0
##   Morus           0     0           0      0          0        0      0
##   Olea            0     0           0      0          0        0      0
##   Phildelphus     0     0           0      0          0        0      0
##   Populus         0     0           0      0          0        0      0
##   Prunus          0     0           0      0          0        0      0
##   Pterocarya      0     0           0      0          0        0      0
##   Quercus         1     1           0      0          0        0      0
##   Rhododendron    0     0           0      0          0        0      0
##   Salix           0     0           0      0          0        0      0
##   Sorbus          0     0           0      0          0        0      0
##   Tilia           0     0           0      0          0        0      0
##   Ulmus           0     0           0      0          0        0      0
##   Viburnum        0     0           0      0          0        0      0
##   Zelkova         0     0           0      0          0        0      0
##               Reference
## Prediction     Cercis Cornus Cotinus Crataegus Cytisus Eucalyptus Fagus
##   Acer              0      0       0         0       0          0     0
##   Alnus             0      0       0         0       0          0     0
##   Arundinaria       0      0       0         0       0          0     0
##   Betula            0      0       0         0       0          0     0
##   Callicarpa        0      0       0         0       0          0     0
##   Castanea          0      0       0         0       0          0     0
##   Celtis            0      0       0         0       0          0     0
##   Cercis            0      0       0         0       0          0     0
##   Cornus            0      2       0         0       0          0     0
##   Cotinus           0      0       0         0       0          0     0
##   Crataegus         0      0       0         0       0          0     0
##   Cytisus           0      0       0         0       1          0     0
##   Eucalyptus        0      1       0         0       0          0     0
##   Fagus             0      0       0         0       0          0     0
##   Ginkgo            0      0       0         0       0          0     0
##   Ilex              0      0       0         0       0          0     0
##   Liquidambar       0      0       0         0       0          0     0
##   Liriodendron      0      0       0         0       0          0     0
##   Lithocarpus       0      0       0         0       0          0     0
##   Magnolia          0      0       0         0       0          0     0
##   Morus             0      0       0         0       0          0     0
##   Olea              0      0       0         0       0          0     0
##   Phildelphus       0      0       0         0       0          0     0
##   Populus           0      0       0         0       0          0     0
##   Prunus            0      0       0         0       0          0     0
##   Pterocarya        0      0       0         0       0          0     0
##   Quercus           0      0       0         1       0          0     0
##   Rhododendron      0      0       0         0       0          0     0
##   Salix             0      0       0         0       0          0     0
##   Sorbus            0      0       0         0       0          0     0
##   Tilia             0      0       0         0       0          0     0
##   Ulmus              0      0       0         0       0          0     0
##   Viburnum           0      0       0         0       0          0     0
##   Zelkova            0      0       0         0       0          0     0
##                Reference
## Prediction     Ginkgo Ilex Liquidambar Liriodendron Lithocarpus Magnolia
##   Acer              0    0           0            0           0        0
##   Alnus             0    0           0            0           0        0
##   Arundinaria       0    0           0            0           0        0
##   Betula            0    0           0            0           0        0
##   Callicarpa        0    0           0            0           0        0
##   Castanea          0    0           0            0           0        0
##   Celtis            0    0           0            0           0        0
##   Cercis            0    0           0            0           0        0
##   Cornus            0    0           0            0           0        0
##   Cotinus           0    0           0            0           0        0
##   Crataegus         0    0           0            0           0        0
##   Cytisus           0    0           0            0           0        0
##   Eucalyptus        0    0           0            0           0        0
##   Fagus             0    0           0            0           0        0
##   Ginkgo            2    0           0            0           0        0
##   Ilex              0    2           0            0           0        0
##   Liquidambar       0    0           3            0           0        0
##   Liriodendron      0    1           0            3           0        0
##   Lithocarpus       0    0           0            0           3        0
##   Magnolia          0    0           0            0           2        1
##   Morus             0    0           0            0           0        0
##   Olea              0    0           0            0           0        0
##   Phildelphus       0    0           0            0           0        0
##   Populus           0    0           0            0           0        0
##   Prunus            0    0           0            0           0        0
##   Pterocarya        0    0           0            0           0        0
##   Quercus           0    0           0            0           0        1
##   Rhododendron      0    0           0            0           0        0
##   Salix             0    0           0            0           0        0
##   Sorbus            0    0           0            0           0        0
##   Tilia             0    0           0            0           0        0
##   Ulmus             0    0           0            0           0        0
##   Viburnum          0    0           0            0           0        0
##   Zelkova           0    0           0            0           0        0
##                Reference
## Prediction     Morus Olea Phildelphus Populus Prunus Pterocarya Quercus
##   Acer             0    0           0       0      0          0      17
##   Alnus            0    0           0       0      0          0      11
##   Arundinaria      0    0           0       0      0          0       4
##   Betula           0    0           0       0      0          0      11
##   Callicarpa       0    0           0       0      0          0       3
##   Castanea         0    0           0       0      0          0       1
##   Celtis           0    0           0       0      0          0       4
##   Cercis           0    0           0       0      0          0       5
##   Cornus           0    0           0       0      0          0       9
##   Cotinus          0    0           0       0      0          0       3
##   Crataegus        0    0           0       0      0          0       2
##   Cytisus          0    0           0       0      0          0       2
##   Eucalyptus       0    0           0       0      0          0       8
##   Fagus            0    0           0       0      0          0       2
##   Ginkgo           0    0           0       0      0          0       3
##   Ilex             0    0           0       0      0          0       6
##   Liquidambar      0    0           0       0      0          0       0
##   Liriodendron     0    0           0       0      0          0       0
##   Lithocarpus      0    0           0       0      0          0      10
##   Magnolia         0    0           0       0      0          0       3
##   Morus            2    0           0       0      0          0       3
##   Olea             0    0           0       0      0          0       1
##   Phildelphus      0    0           1       0      0          0       2
##   Populus          0    0           0       0      0          0      12
##   Prunus           0    0           0       0      0          0       9
##   Pterocarya       0    0           0       0      0          0       5
##   Quercus          0    1           1       0      0          0     110
##   Rhododendron     0    0           0       0      0          0       2
##   Salix            0    0           0       0      0          0       3
##   Sorbus           0    0           0       0      0          0       2
##   Tilia            0    0           0       0      0          0       5
##   Ulmus            0    0           0       0      0          0       2
##   Viburnum         0    0           0       0      0          0       6
##   Zelkova          0    0           0       0      0          0       3
##               Reference
## Prediction     Rhododendron Salix Sorbus Tilia Ulmus Viburnum Zelkova
##   Acer                    0     0      0     0     0        0       0
##   Alnus                   0     0      0     0     0        0       0
##   Arundinaria             0     0      0     0     0        0       0
##   Betula                  0     0      0     0     0        0       0
##   Callicarpa              0     0      0     0     0        0       0
##   Castanea                0     0      0     0     1        0       0
##   Celtis                  0     0      0     0     0        0       0
##   Cercis                  0     0      0     0     0        0       0
##   Cornus                  0     0      0     1     0        0       0
##   Cotinus                 0     0      0     0     0        0       0
##   Crataegus               0     0      0     0     0        0       0
##   Cytisus                 0     0      0     0     0        0       0
##   Eucalyptus              0     0      0     0     0        0       0
##   Fagus                   0     0      0     0     0        0       0
##   Ginkgo                  0     0      0     0     0        0       0
##   Ilex                    0     0      0     0     0        0       0
##   Liquidambar             0     0      0     0     0        0       0
##   Liriodendron            0     0      0     0     0        0       0
##   Lithocarpus             0     0      0     0     0        0       0
##   Magnolia                0     0      0     0     0        0       0
##   Morus                   0     0      0     0     0        0       0
##   Olea                    0     0      0     0     0        0       0
##   Phildelphus             0     0      0     0     0        0       0
##   Populus                 0     0      0     0     0        0       0
##   Prunus                  0     0      0     0     0        0       0
##   Pterocarya              0     0      0     0     0        0       0
##   Quercus                 3     2      0     1     0        0       0
##   Rhododendron            2     0      0     0     0        0       0
##   Salix                   0     1      0     0     0        0       0
##   Sorbus                  0     0      0     1     0        0       0
##   Tilia                   0     0      1     5     0        0       0
##   Ulmus                   0     0      0     0     3        0       0
##   Viburnum                0     0      0     0     0        0       0
##   Zelkova                 0     0      0     0     0        0       1
##
## Overall Statistics
##
##                Accuracy : 0.4784
##                  95% CI : (0.4248, 0.5324)
##     No Information Rate : 0.7752
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.2741
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Acer Class: Alnus Class: Arundinaria
## Sensitivity              0.82353      0.90000                 NA
## Specificity              0.94848      0.96736            0.98847
## Pos Pred Value           0.45161      0.45000                 NA
## Neg Pred Value           0.99051      0.99694                 NA
## Prevalence               0.04899      0.02882            0.00000
## Detection Rate           0.04035      0.02594            0.00000
## Detection Prevalence     0.08934      0.05764            0.01153
## Balanced Accuracy        0.88601      0.93368                 NA
##                      Class: Betula Class: Callicarpa Class: Castanea
## Sensitivity                     NA                NA        1.000000
## Specificity                 0.9683          0.991354        0.994220
## Pos Pred Value                  NA                NA        0.333333
## Neg Pred Value                  NA                NA        1.000000
## Prevalence                  0.0000          0.000000        0.002882
## Detection Rate              0.0000          0.000000        0.002882
## Detection Prevalence        0.0317          0.008646        0.008646
## Balanced Accuracy               NA                NA        0.997110
##                      Class: Celtis Class: Cercis Class: Cornus
## Sensitivity                     NA            NA      0.666667
## Specificity                0.98847       0.98559      0.970930
## Pos Pred Value                  NA            NA      0.166667
## Neg Pred Value                  NA            NA      0.997015
## Prevalence                 0.00000       0.00000      0.008646
## Detection Rate             0.00000       0.00000      0.005764
## Detection Prevalence       0.01153       0.01441      0.034582
## Balanced Accuracy               NA            NA      0.818798
##                      Class: Cotinus Class: Crataegus Class: Cytisus
## Sensitivity                      NA         0.000000       1.000000
## Specificity                0.991354         0.994220       0.994220
## Pos Pred Value                   NA         0.000000       0.333333
## Neg Pred Value                   NA         0.997101       1.000000
## Prevalence                 0.000000         0.002882       0.002882
## Detection Rate             0.000000         0.000000       0.002882
## Detection Prevalence       0.008646         0.005764       0.008646
## Balanced Accuracy                NA         0.497110       0.997110
##                      Class: Eucalyptus Class: Fagus Class: Ginkgo
## Sensitivity                         NA           NA      1.000000
## Specificity                    0.97118     0.994236      0.988406
## Pos Pred Value                      NA           NA      0.333333
## Neg Pred Value                      NA           NA      1.000000
## Prevalence                     0.00000     0.000000      0.005764
## Detection Rate                 0.00000     0.000000      0.005764
## Detection Prevalence           0.02882     0.005764      0.017291
## Balanced Accuracy                   NA           NA      0.994203
##                      Class: Ilex Class: Liquidambar Class: Liriodendron
## Sensitivity             0.666667           1.000000            1.000000
## Specificity             0.982558           1.000000            0.997093
## Pos Pred Value          0.250000           1.000000            0.750000
## Neg Pred Value          0.997050           1.000000            1.000000
## Prevalence              0.008646           0.008646            0.008646
## Detection Rate          0.005764           0.008646            0.008646
## Detection Prevalence    0.023055           0.008646            0.011527
## Balanced Accuracy       0.824612           1.000000            0.998547
##                      Class: Lithocarpus Class: Magnolia Class: Morus
## Sensitivity                    0.600000        0.500000     1.000000
## Specificity                    0.970760        0.985507     0.991304
## Pos Pred Value                 0.230769        0.166667     0.400000
## Neg Pred Value                 0.994012        0.997067     1.000000
## Prevalence                     0.014409        0.005764     0.005764
## Detection Rate                 0.008646        0.002882     0.005764
## Detection Prevalence           0.037464        0.017291     0.014409
## Balanced Accuracy              0.785380        0.742754     0.995652
##                      Class: Olea Class: Phildelphus Class: Populus
## Sensitivity             0.000000           0.500000             NA
## Specificity             0.997110           0.994203        0.96542
## Pos Pred Value          0.000000           0.333333             NA
## Neg Pred Value          0.997110           0.997093             NA
## Prevalence              0.002882           0.005764        0.00000
## Detection Rate          0.000000           0.002882        0.00000
## Detection Prevalence    0.002882           0.008646        0.03458
## Balanced Accuracy       0.498555           0.747101             NA
##                      Class: Prunus Class: Pterocarya Class: Quercus
## Sensitivity                     NA                NA         0.4089
## Specificity                0.97406           0.98559         0.8462
## Pos Pred Value                  NA                NA         0.9016
## Neg Pred Value                  NA                NA         0.2933
## Prevalence                 0.00000           0.00000         0.7752
## Detection Rate             0.00000           0.00000         0.3170
## Detection Prevalence       0.02594           0.01441         0.3516
## Balanced Accuracy               NA                NA         0.6275
##                      Class: Rhododendron Class: Salix Class: Sorbus
## Sensitivity                     0.400000     0.333333      0.000000
## Specificity                     0.994152     0.991279      0.991329
## Pos Pred Value                  0.500000     0.250000      0.000000
## Neg Pred Value                  0.991254     0.994169      0.997093
## Prevalence                      0.014409     0.008646      0.002882
## Detection Rate                  0.005764     0.002882      0.000000
## Detection Prevalence            0.011527     0.011527      0.008646
## Balanced Accuracy               0.697076     0.662306      0.495665
##                      Class: Tilia Class: Ulmus Class: Viburnum
## Sensitivity              0.62500    0.750000            NA
## Specificity              0.98230    0.994169       0.98271
## Pos Pred Value           0.45455    0.600000            NA
## Neg Pred Value           0.99107    0.997076            NA
## Prevalence               0.02305    0.011527       0.00000
## Detection Rate           0.01441    0.008646       0.00000
## Detection Prevalence     0.03170    0.014409       0.01729
## Balanced Accuracy        0.80365    0.872085            NA
##                        Class: Zelkova
## Sensitivity                  1.000000
## Specificity                  0.991329
## Pos Pred Value               0.250000
## Neg Pred Value               1.000000
## Prevalence                   0.002882
## Detection Rate               0.002882
## Detection Prevalence         0.011527
## Balanced Accuracy            0.995665
```

PART 2 ANSWER We see shades of purple on this plot representing the different genus' from the model. There are no straight lines which is an indicator of decision gradients. It is more likely to be predicted a certain genus in the dark purple, but it is not a guarantee that the genus will be what is predicted.

##(e) (15 points) This question will use a non-linear kernel for the SVM and compare results.

(i) Modify your function in part (c) to find the optimal cost value for the SVM on the **training data** with **radial kernel** with gamma parameter 0.55. Use the same cost range. Report the optimal cost.

(ii) Run the radial SVM model with these optimal parameters on the training data.

(iii) Repeat part (d)(iii) but for the radial SVM model instead of the linear one.

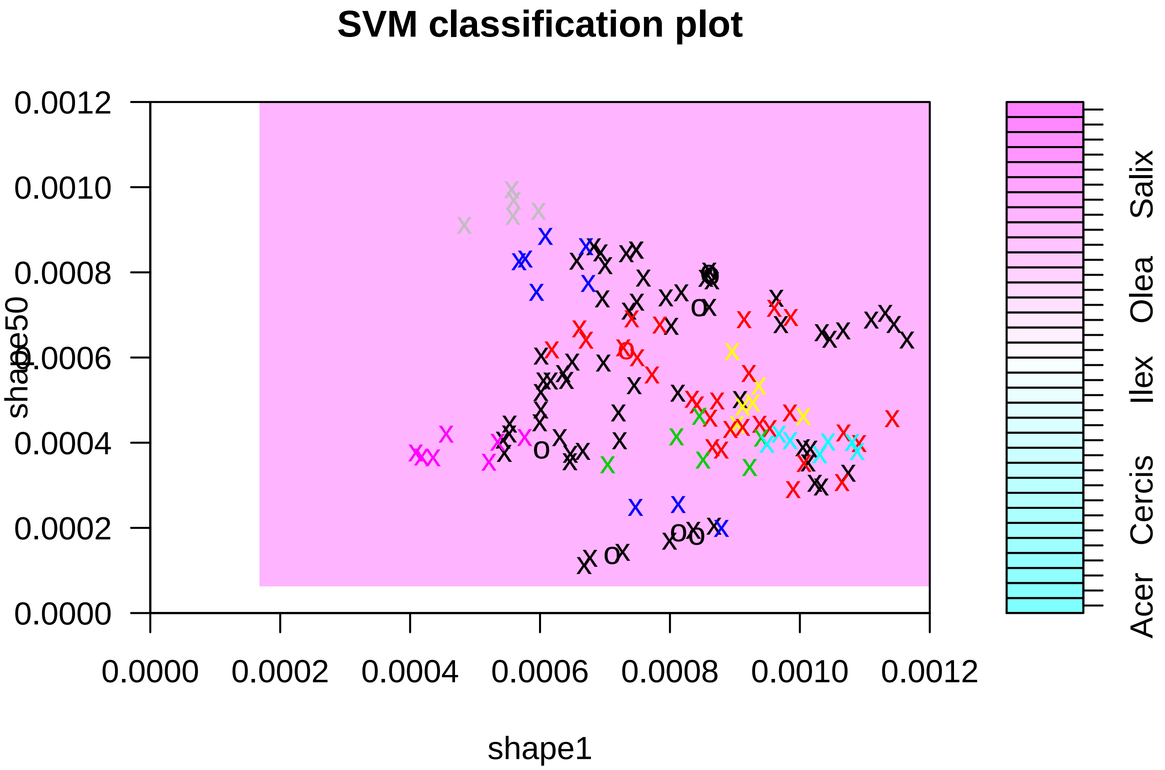
(iv) Discuss briefly your results in (e)(iii) as compared to (d)(iii) **using concepts discussed in class**.

```
cost_out <- seq(from = 0.1, to =5.1, by = 1)
missed = rep(NA, length(cost_out))
for(i in 1:length(cost_out)){
  svm.mod = svm(formula = genus~., data = select(leaf_train, -c(id, species)), kernel = "radial", cost =
  missed[i] = (sum(svm.mod$fitted != leaf_train$genus))/nrow(leaf_train)
}

data.frame(cbind(cost_out, missed))
```

```
##   cost_out     missed
## 1      0.1 0.59875583
## 2      1.1 0.07620529
## 3      2.1 0.01244168
## 4      3.1 0.00311042
## 5      4.1 0.00000000
## 6      5.1 0.00000000
```

PART 3 ANSWER c has a best value of 4.1 and 5.1. For continuity I chose to use c as 5.1 in this model.

```
svm.mod = svm(formula = genus~., data = select(leaf_train, -c(id, species)), kernel = "radial", cost =

plot(svm.mod, data = select(leaf_train, -c(id, species)), shape50~shape1, xlim = c(0, 0.0012), ylim = c
```



```r
svm.pred = predict(svm.mod, newdata = leaf_test)
library(caret)
confusion = confusionMatrix(leaf_test$genus, svm.pred)
confusion
```

```
## Confusion Matrix and Statistics
##
##                Reference
## Prediction     Acer Alnus Arundinaria Betula Callicarpa Castanea Celtis
##   Acer           22     0           0      0          0        0      0
##   Alnus           1    11           0      0          1        0      0
##   Arundinaria     0     0           0      0          0        0      0
##   Betula          0     0           0      0          0        0      0
##   Callicarpa      1     1           0      0          1        0      0
##   Castanea        0     0           0      0          0        0      0
##   Celtis          0     0           0      0          0        0      0
##   Cercis          0     0           0      0          0        0      0
##   Cornus          1     0           0      0          0        0      0
##   Cotinus         0     0           0      0          0        0      0
##   Crataegus       0     0           0      1          0        0      0
##   Cytisus         0     0           0      0          0        0      0
##   Eucalyptus      0     0           0      0          0        0      0
##   Fagus           0     0           0      0          0        0      0
##   Ginkgo          0     0           0      0          0        0      0
##   Ilex            0     0           0      0          0        0      0
##   Liquidambar     0     0           0      0          0        0      0
##   Liriodendron    0     0           0      0          0        0      0
##   Lithocarpus     0     0           0      0          0        0      0
##   Magnolia        0     0           0      0          0        0      0
##   Morus           0     0           0      0          0        0      0
##   Olea             0     0           0      0          0        0      0
##   Phildelphus      0     0           0      0          0        0      0
##   Populus          1     0           0      0          0        0      0
##   Prunus           0     0           0      0          0        0      0
##   Pterocarya       1     0           0      0          0        0      0
##   Quercus          0     1           0      1          0        0      0
##   Rhododendron     0     0           0      0          0        0      0
##   Salix            0     0           0      0          0        0      0
##   Sorbus           0     0           0      0          0        0      0
##   Tilia            0     0           0      0          0        0      0
##   Ulmus            0     0           0      0          0        0      0
##   Viburnum         0     0           0      0          0        0      0
##   Zelkova          1     0           0      0          0        0      0
##               Reference
## Prediction     Cercis Cornus Cotinus Crataegus Cytisus Eucalyptus Fagus
##   Acer              0      0       0         0       0          0     0
##   Alnus             0      1       0         0       0          0     0
##   Arundinaria       0      0       0         0       0          0     0
##   Betula            0      0       0         1       0          0     0
##   Callicarpa        0      0       0         0       0          0     0
##   Castanea          0      0       0         0       0          0     0
##   Celtis            0      0       0         0       0          0     0
##   Cercis            1      0       0         0       0          0     0
##   Cornus            0      5       0         0       0          0     0
##   Cotinus           0      0       1         0       0          0     0
##   Crataegus         0      0       0         0       0          0     0
##   Cytisus           0      0       0         0       2          0     0
##   Eucalyptus        0      0       0         0       0          6     0
##   Fagus             0      0       0         1       0          0     1
##   Ginkgo            0      0       0         0       0          0     0
##   Ilex              0      0       0         0       0          0     0
##   Liquidambar       0      0       0         0       0          0     0
##   Liriodendron      0      0       0         0       0          0     0
##   Lithocarpus       0      0       0         0       0          0     0
##   Magnolia          0      0       0         0       0          0     0
##   Morus             0      0       0         0       0          0     0
##   Olea              0      0       0         0       0          0     0
##   Phildelphus       0      0       0         0       0          0     0
##   Populus           0      0       0         0       0          0     0
##   Prunus            0      0       0         0       0          0     0
##   Pterocarya        0      0       0         0       0          0     0
##   Quercus           1      0       0         0       1          3     0
##   Rhododendron      0      0       0         0       0          0     0
##   Salix             0      0       0         0       0          0     0
##   Sorbus            0      0       0         0       0          0     0
##   Tilia             0      0       0         0       0          0     0
##   Ulmus             0      0       0         0       0          0     0
##   Viburnum          0      0       0         0       0          0     0
##   Zelkova           0      0       0         0       0          0     0
##               Reference
## Prediction     Ginkgo Ilex Liquidambar Liriodendron Lithocarpus Magnolia
##   Acer              0    0           0            0           0        0
##   Alnus             0    0           0            0           0        0
##   Arundinaria       0    0           0            0           0        0
##   Betula            0    0           0            0           0        0
##   Callicarpa        0    0           0            0           0        0
##   Castanea          0    0           0            0           0        0
##   Celtis            0    0           0            0           0        0
##   Cercis            0    0           0            0           0        0
##   Cornus            0    0           0            0           0        0
##   Cotinus           0    0           0            0           0        0
##   Crataegus         0    0           0            0           0        0
##   Cytisus           0    0           0            0           0        0
##   Eucalyptus        0    0           0            0           0        0
##   Fagus             0    0           0            0           0        0
##   Ginkgo            3    0           0            0           0        0
##   Ilex              0    6           0            0           0        0
##   Liquidambar       0    0           0            0           0        0
##   Liriodendron      0    0           0            1           0        0
##   Lithocarpus       0    0           0            0           1        5
##   Magnolia          0    0           0            0           0        2
##   Morus             0    0           0            0           0        0
##   Olea              0    0           0            0           0        0
##   Phildelphus       0    0           0            0           0        0
##   Populus           0    0           0            0           0        0
##   Prunus            0    0           0            0           0        0
##   Pterocarya        0    0           0            0           0        0
##   Quercus           0    0           0            0           0        0
##   Rhododendron      0    0           0            0           0        0
##   Salix             0    0           0            0           0        0
##   Sorbus            0    0           0            0           0        0
##   Tilia             0    0           0            0           0        0
##   Ulmus             0    0           0            0           0        0
##   Viburnum          0    0           0            0           0        0
##   Zelkova           0    0           0            0           0        0
##                Reference
## Prediction      Morus Olea Phildelphus Populus Prunus Pterocarya Quercus
##   Acer              0    0           0       0      0          0       9
##   Alnus             0    0           0       0      0          0       6
##   Arundinaria       0    0           0       0      0          0       4
##   Betula            0    0           0       0      0          0      10
##   Callicarpa        0    0           0       0      0          0       0
##   Castanea          0    0           0       0      0          0       3
##   Celtis            0    0           0       0      0          0       4
##   Cercis            0    0           0       0      0          0       4
##   Cornus            0    0           0       0      0          0       6
##   Cotinus           0    0           0       0      0          0       2
##   Crataegus         0    1           0       0      0          0       0
##   Cytisus           0    0           0       0      0          0       0
##   Eucalyptus        0    0           0       0      0          0       4
##   Fagus             0    0           0       0      0          0       0
##   Ginkgo            0    0           0       0      0          0       3
##   Ilex              0    0           0       0      0          0       2
##   Liquidambar       0    0           0       0      0          0       3
##   Liriodendron      0    0           0       0      0          0       3
##   Lithocarpus       0    0           0       0      0          0       7
##   Magnolia          0    0           0       0      0          0       4
##   Morus             5    0           0       0      0          0       0
##   Olea             0    1           0       0      0          0       0
##   Phildelphus      0    0           2       0      0          0       1
##   Populus          0    1           1       3      0          0       6
##   Prunus           0    0           0       0      7          0       2
##   Pterocarya       0    0           0       0      0          1       3
##   Quercus          0    0           0       0      0          0     115
##   Rhododendron     0    0           0       0      0          0       4
##   Salix            0    0           0       0      0          0       3
##   Sorbus           0    0           0       0      0          0       2
##   Tilia            0    0           0       0      0          0       3
##   Ulmus            0    0           0       0      0          0       0
##   Viburnum         0    0           0       0      0          0       4
##   Zelkova          0    0           0       0      0          0       3
##               Reference
## Prediction     Rhododendron Salix Sorbus Tilia Ulmus Viburnum Zelkova
##   Acer                    0     0      0     0     0        0       0
##   Alnus                   0     0      0     0     0        0       0
##   Arundinaria             0     0      0     0     0        0       0
##   Betula                  0     0      0     0     0        0       0
##   Callicarpa              0     0      0     0     0        0       0
##   Castanea                0     0      0     0     0        0       0
##   Celtis                  0     0      0     0     0        0       0
##   Cercis                  0     0      0     0     0        0       0
##   Cornus                  0     0      0     0     0        0       0
##   Cotinus                 0     0      0     0     0        0       0
##   Crataegus               0     0      0     0     0        0       0
##   Cytisus                 0     0      0     0     0        1       0
##   Eucalyptus              0     0      0     0     0        0       0
##   Fagus                   0     0      0     0     0        0       0
##   Ginkgo                  0     0      0     0     0        0       0
##   Ilex                    0     0      0     0     0        0       0
##   Liquidambar             0     0      0     0     0        0       0
##   Liriodendron            0     0      0     0     0        0       0
##   Lithocarpus             0     0      0     0     0        0       0
##   Magnolia                0     0      0     0     0        0       0
##   Morus                   0     0      0     0     0        0       0
##   Olea                    0     0      0     0     0        0       0
##   Phildelphus             0     0      0     0     0        0       0
##   Populus                 0     0      0     0     0        0       0
##   Prunus                  0     0      0     0     0        0       0
##   Pterocarya              0     0      0     0     0        0       0
##   Quercus                 0     0      0     0     0        0       0
##   Rhododendron            0     0      0     0     0        0       0
##   Salix                   0     1      0     0     0        0       0
##   Sorbus                  0     0      0     1     0        0       0
##   Tilia                   0     0      0     7     1        0       0
##   Ulmus                   0     0      0     0     5        0       0
##   Viburnum                0     0      0     0     0        2       0
##   Zelkova                 0     0      0     0     0        0       0
##
## Overall Statistics
##
##                Accuracy : 0.611
##                  95% CI : (0.5574, 0.6625)
##     No Information Rate : 0.634
##     P-Value [Acc > NIR] : 0.8284
##
##                   Kappa : 0.4898
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Acer Class: Alnus Class: Arundinaria
## Sensitivity              0.78571      0.84615                 NA
## Specificity              0.97179      0.97305            0.98847
## Pos Pred Value           0.70968      0.55000                 NA
## Neg Pred Value           0.98101      0.99388                 NA
## Prevalence               0.08069      0.03746            0.00000
## Detection Rate           0.06340      0.03170            0.00000
## Detection Prevalence     0.08934      0.05764            0.01153
## Balanced Accuracy        0.87875      0.90960                 NA
##                      Class: Betula Class: Callicarpa Class: Castanea
## Sensitivity               0.000000          0.500000              NA
## Specificity               0.968116          0.994203        0.991354
## Pos Pred Value            0.000000          0.333333              NA
## Neg Pred Value            0.994048          0.997093              NA
## Prevalence                0.005764          0.005764        0.000000
## Detection Rate            0.000000          0.002882        0.000000
## Detection Prevalence      0.031700          0.008646        0.008646
## Balanced Accuracy         0.484058          0.747101              NA
##                      Class: Celtis Class: Cercis Class: Cornus
## Sensitivity                     NA      0.500000       0.83333
## Specificity                0.98847      0.988406       0.97947
## Pos Pred Value                  NA      0.200000       0.41667
## Neg Pred Value                  NA      0.997076       0.99701
## Prevalence                 0.00000      0.005764       0.01729
## Detection Rate             0.00000      0.002882       0.01441
## Detection Prevalence       0.01153      0.014409       0.03458
## Balanced Accuracy               NA      0.744203       0.90640
##                      Class: Cotinus Class: Crataegus Class: Cytisus
## Sensitivity                1.000000         0.000000       0.666667
## Specificity                0.994220         0.994203       0.997093
## Pos Pred Value             0.333333         0.000000       0.666667
## Neg Pred Value             1.000000         0.994203       0.997093
## Prevalence                 0.002882         0.005764       0.008646
## Detection Rate             0.002882         0.000000       0.005764
## Detection Prevalence       0.008646         0.005764       0.008646
## Balanced Accuracy          0.997110         0.497101       0.831880
##                      Class: Eucalyptus Class: Fagus Class: Ginkgo
## Sensitivity                    0.66667     1.000000      1.000000
## Specificity                    0.98817     0.997110      0.991279
## Pos Pred Value                 0.60000     0.500000      0.500000
## Neg Pred Value                 0.99110     1.000000      1.000000
## Prevalence                     0.02594     0.002882      0.008646
## Detection Rate                 0.01729     0.002882      0.008646
## Detection Prevalence           0.02882     0.005764      0.017291
## Balanced Accuracy              0.82742     0.998555      0.995640
##                      Class: Ilex Class: Liquidambar Class: Liriodendron
## Sensitivity              1.00000                 NA            1.000000
## Specificity              0.99413           0.991354            0.991329
## Pos Pred Value           0.75000                 NA            0.250000
## Neg Pred Value           1.00000                 NA            1.000000
## Prevalence               0.01729           0.000000            0.002882
## Detection Rate           0.01729           0.000000            0.002882
## Detection Prevalence     0.02305           0.008646            0.011527
## Balanced Accuracy        0.99707                 NA            0.995665
##                      Class: Lithocarpus Class: Magnolia Class: Morus
## Sensitivity                    1.000000        0.285714      1.00000
## Specificity                    0.965318        0.988235      1.00000
## Pos Pred Value                 0.076923        0.333333      1.00000
## Neg Pred Value                 1.000000        0.985337      1.00000
## Prevalence                     0.002882        0.020173      0.01441
## Detection Rate                 0.002882        0.005764      0.01441
## Detection Prevalence           0.037464        0.017291      0.01441
## Balanced Accuracy              0.982659        0.636975      1.00000
##                      Class: Olea Class: Phildelphus Class: Populus
## Sensitivity             0.333333           0.666667       1.000000
## Specificity             1.000000           0.997093       0.973837
## Pos Pred Value          1.000000           0.666667       0.250000
## Neg Pred Value          0.994220           0.997093       1.000000
## Prevalence              0.008646           0.008646       0.008646
## Detection Rate          0.002882           0.005764       0.008646
## Detection Prevalence    0.002882           0.008646       0.034582
## Balanced Accuracy       0.666667           0.831880       0.986919
##                      Class: Prunus Class: Pterocarya Class: Quercus
## Sensitivity                1.00000          1.000000         0.5227
## Specificity                0.99412          0.988439         0.9449
## Pos Pred Value             0.77778          0.200000         0.9426
## Neg Pred Value             1.00000          1.000000         0.5333
## Prevalence                 0.02017          0.002882         0.6340
## Detection Rate             0.02017          0.002882         0.3314
## Detection Prevalence       0.02594          0.014409         0.3516
## Balanced Accuracy          0.99706          0.994220         0.7338
##                      Class: Rhododendron Class: Salix Class: Sorbus
## Sensitivity                           NA     1.000000            NA
## Specificity                      0.98847     0.991329      0.991354
## Pos Pred Value                        NA     0.250000            NA
## Neg Pred Value                        NA     1.000000            NA
## Prevalence                       0.00000     0.002882      0.000000
## Detection Rate                   0.00000     0.002882      0.000000
## Detection Prevalence             0.01153     0.011527      0.008646
## Balanced Accuracy                     NA     0.995665            NA
##                      Class: Tilia Class: Ulmus Class: Viburnum
## Sensitivity               0.87500      0.83333        0.666667
## Specificity               0.98820      1.00000        0.988372
## Pos Pred Value            0.63636      1.00000        0.333333
## Neg Pred Value            0.99702      0.99708        0.997067
## Prevalence                0.02305      0.01729        0.008646
## Detection Rate            0.02017      0.01441        0.005764
## Detection Prevalence      0.03170      0.01441        0.017291
## Balanced Accuracy         0.93160      0.91667        0.827519
##                      Class: Zelkova
## Sensitivity                      NA
## Specificity                 0.98847
## Pos Pred Value                   NA
## Neg Pred Value                   NA
## Prevalence                  0.00000
## Detection Rate              0.00000
## Detection Prevalence        0.01153
## Balanced Accuracy                NA
```

PART 4 ANSWER The non-linear SVM model has better accuracy (.61 vs .47). This can also be seen with better specificity and sensitivity scores.