Multinomial Regression

B. Whillock

May 16, 2019

library(tidyverse)  
# for data manipulation and presentation  
  
library(rsample)  
# for stratified sampling on classification problems  
  
library(nnet)  
# for multinomial regression  
  
library(caret)  
# for cross validation

## Iris Data

# the dataset iris is native to base R  
# convert to tibble for convenience  
iris <- as\_tibble(iris)  
iris

## # A tibble: 150 x 5  
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## <dbl> <dbl> <dbl> <dbl> <fct>   
## 1 5.1 3.5 1.4 0.2 setosa   
## 2 4.9 3 1.4 0.2 setosa   
## 3 4.7 3.2 1.3 0.2 setosa   
## 4 4.6 3.1 1.5 0.2 setosa   
## 5 5 3.6 1.4 0.2 setosa   
## 6 5.4 3.9 1.7 0.4 setosa   
## 7 4.6 3.4 1.4 0.3 setosa   
## 8 5 3.4 1.5 0.2 setosa   
## 9 4.4 2.9 1.4 0.2 setosa   
## 10 4.9 3.1 1.5 0.1 setosa   
## # ... with 140 more rows

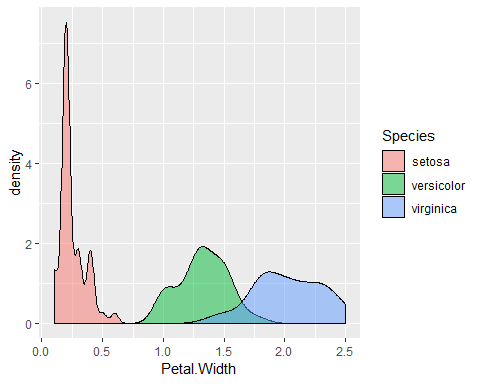
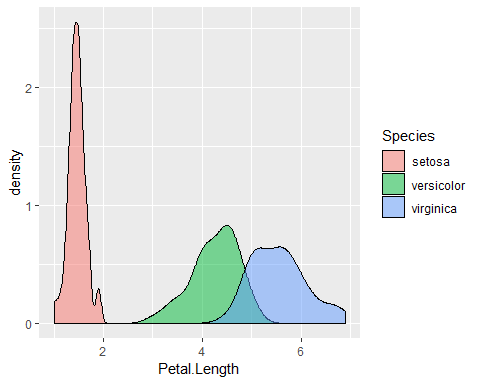
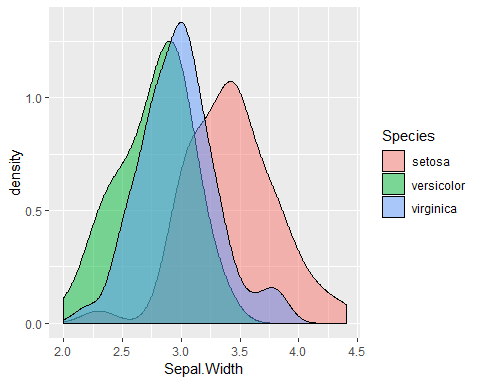
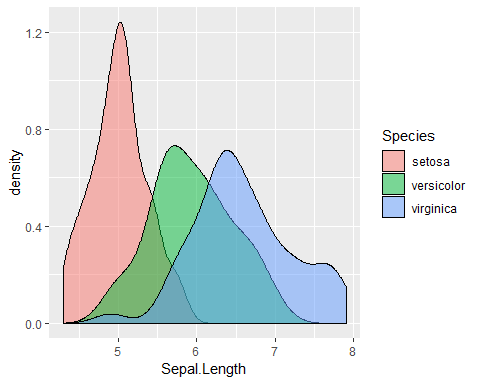
summary(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
## Species   
## setosa :50   
## versicolor:50   
## virginica :50   
##   
##   
##

dim(iris)

## [1] 150 5

## Density plots of variables by their class:



## Training

# 50/50 train-test split with near equal distribution of the dependent variable  
set.seed(123)  
split = initial\_split(iris, prop=.5, strata="Species")  
train = training(split)  
test = testing(split)

# verify distribution of the dependent variable  
table(train$Species) %>% prop.table()

##   
## setosa versicolor virginica   
## 0.3333333 0.3333333 0.3333333

table(test$Species) %>% prop.table()

##   
## setosa versicolor virginica   
## 0.3333333 0.3333333 0.3333333

# fit the training data with multinomial regression   
fit = multinom(Species ~ ., data=train, maxit=10000, trace = FALSE)

summary(fit)

## Call:  
## multinom(formula = Species ~ ., data = train, maxit = 10000,   
## trace = FALSE)  
##   
## Coefficients:  
## (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width  
## versicolor 80.66450 1.165827 -37.42907 8.812377 2.678057  
## virginica -92.97015 -33.091202 -89.60040 102.418118 48.633214  
##   
## Std. Errors:  
## (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width  
## versicolor 30755.14 3212.989 12098.58 6328.777 11809.83  
## virginica 30903.10 3395.940 12143.94 6464.139 11872.64  
##   
## Residual Deviance: 0.0001998204   
## AIC: 20.0002

confusionMatrix(predict(fit, test), test$Species)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction setosa versicolor virginica  
## setosa 24 0 0  
## versicolor 1 23 2  
## virginica 0 2 23  
##   
## Overall Statistics  
##   
## Accuracy : 0.9333   
## 95% CI : (0.8512, 0.978)  
## No Information Rate : 0.3333   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: setosa Class: versicolor Class: virginica  
## Sensitivity 0.9600 0.9200 0.9200  
## Specificity 1.0000 0.9400 0.9600  
## Pos Pred Value 1.0000 0.8846 0.9200  
## Neg Pred Value 0.9804 0.9592 0.9600  
## Prevalence 0.3333 0.3333 0.3333  
## Detection Rate 0.3200 0.3067 0.3067  
## Detection Prevalence 0.3200 0.3467 0.3333  
## Balanced Accuracy 0.9800 0.9300 0.9400

## Training with Cross Validation

# Set up for 10 fold CV and hyperparameter tuning  
train\_control <- trainControl(  
 method = "cv",  
 number = 10  
)  
tune\_grid <- expand.grid(  
 decay = c(1,.5,.1, 0)  
)

# fit the training data with multinomial regression, 10-fold CV,  
# and hyperparameter tuning  
fit2 <- train(  
 data=train,  
 Species ~ .,  
 method = "multinom",  
 maxit = 10000,  
 trControl = train\_control,  
 tuneGrid = tune\_grid,  
 trace = FALSE  
)

fit2

## Penalized Multinomial Regression   
##   
## 75 samples  
## 4 predictor  
## 3 classes: 'setosa', 'versicolor', 'virginica'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 67, 68, 67, 68, 68, 68, ...   
## Resampling results across tuning parameters:  
##   
## decay Accuracy Kappa   
## 0.0 0.9763889 0.9647287  
## 0.1 0.9763889 0.9647287  
## 0.5 0.9621032 0.9428537  
## 1.0 0.9603175 0.9395833  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was decay = 0.1.

summary(fit2)

## Call:  
## nnet::multinom(formula = .outcome ~ ., data = dat, decay = param$decay,   
## maxit = 10000, trace = FALSE)  
##   
## Coefficients:  
## (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width  
## versicolor 1.109127 -0.01583202 -2.107954 1.996663 0.006199233  
## virginica -2.150239 -2.10145097 -3.490157 4.970951 3.435953629  
##   
## Std. Errors:  
## (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width  
## versicolor 10.82594 2.982093 3.019285 3.080229 6.296288  
## virginica 11.64927 3.211665 3.435329 3.380947 6.565313  
##   
## Residual Deviance: 32.47585   
## AIC: 52.47585

confusionMatrix(fit2)

## Cross-Validated (10 fold) Confusion Matrix   
##   
## (entries are percentual average cell counts across resamples)  
##   
## Reference  
## Prediction setosa versicolor virginica  
## setosa 33.3 0.0 0.0  
## versicolor 0.0 32.0 1.3  
## virginica 0.0 1.3 32.0  
##   
## Accuracy (average) : 0.9733

confusionMatrix(predict(fit2, test), test$Species)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction setosa versicolor virginica  
## setosa 25 0 0  
## versicolor 0 24 0  
## virginica 0 1 25  
##   
## Overall Statistics  
##   
## Accuracy : 0.9867   
## 95% CI : (0.9279, 0.9997)  
## No Information Rate : 0.3333   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.98   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: setosa Class: versicolor Class: virginica  
## Sensitivity 1.0000 0.9600 1.0000  
## Specificity 1.0000 1.0000 0.9800  
## Pos Pred Value 1.0000 1.0000 0.9615  
## Neg Pred Value 1.0000 0.9804 1.0000  
## Prevalence 0.3333 0.3333 0.3333  
## Detection Rate 0.3333 0.3200 0.3333  
## Detection Prevalence 0.3333 0.3200 0.3467  
## Balanced Accuracy 1.0000 0.9800 0.9900