Problemas Propuesto Regresión Multinomial

Francisco Javier Mercader Martínez

Problema 1

El dichero **iris** de R, contiene los datos correspondientes de los pétalos y sépalos de tres variedades de flor de iris (setosa, virginica y versicolor).

Se desea realizar un análisis de Regresión Multinomial con el fin de predecir la variedad de la flor en función de las magnitudes de sus pétalos y sépalos. Se pide:

1) Recuperar los datos y realizar un estudio descriptivo previo atendiendo a nuestro objetivo.

```
iris <- data.frame(iris)
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
##
##
##
```

2) Dividir el conjunto de datos en entrenamiento y prueba (70% entrenamiento, 30% prueba). Tomar semilla 123.

```
set.seed(123)

split <- sample(1:nrow(iris), 0.7 * nrow(iris))

# Conjunto de entrenamiento
train_data <- iris[split, ]

# Conjunto de prueba
test_data <- iris[-split, ]
```

3) Con los datos de entrenamiento, obtener el modelo ajustado de Regresión Multinomial usando todos los predictores.

```
library(nnet)
modelo_ajustado <- multinom(Species ~ ., data = train_data)
```

```
## # weights:  18 (10 variable)
## initial  value 115.354290
## iter  10 value 14.037979
## iter  20 value 3.342288
## iter  30 value 2.503699
## iter  40 value 2.171547
## iter  50 value 2.099460
## iter  60 value 1.828506
## iter  70 value 0.904367
## iter  80 value 0.669147
## iter  90 value 0.622003
## iter 100 value 0.609416
## final  value 0.609416
## stopped after 100 iterations
```

```
summary(modelo_ajustado)
```

```
## Call:
## multinom(formula = Species ~ ., data = train_data)
##
## Coefficients:
##            (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width
## versicolor     63.7972    -27.80712   -27.99961      71.5816    18.78823
## virginica    -107.2881    -56.45906   -61.59348     140.6447    82.34126
##
## Std. Errors:
##            (Intercept) Sepal.Length Sepal.Width Petal.Length Petal.Width
## versicolor    119.5758     41.53559    29.48294     45.30698    30.24145
## virginica     119.5759     41.53544    29.48285     45.30703    30.24145
##
## Residual Deviance: 1.218832
## AIC: 21.21883
```

4) Con los datos de entrenamiento, aplicar los métodos de selección de regresores para comprobar si el modelo completo es reducible.

```
modelo_nulo <- multinom(Species ~ 1, data = train_data)
```

```
## # weights:  6 (2 variable)
## initial  value 115.354290
## final  value 115.151643
## converged
```

```
modelo_backward <- step(modelo_ajustado, direction = "backward")
```

```
## Start:  AIC=21.22
## Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width
##
## trying - Sepal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 12.178204
## iter  20 value 5.050112
## iter  30 value 4.842539
## iter  40 value 4.731343
## iter  50 value 4.610612
```

```
## iter  60 value 4.597635
## final  value 4.596501
## converged
## trying - Sepal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
## trying - Petal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 13.208650
## iter  20 value 9.268575
## iter  30 value 8.984624
## iter  40 value 8.955823
## iter  50 value 8.953394
## final  value 8.953389
## converged
## trying - Petal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 17.073901
## iter  20 value 7.858714
## iter  30 value 6.707280
## iter  40 value 6.406272
## iter  50 value 6.381857
## iter  60 value 6.374101
## final  value 6.366160
## converged
##                Df      AIC
## <none>         10 21.21883
## - Sepal.Width   8 24.79663
## - Sepal.Length  8 25.19300
## - Petal.Width   8 28.73232
## - Petal.Length  8 33.90678
```

```
modelo_forward <- step(modelo_nulo, direction = "forward", scope =
 ↪  formula(modelo_ajustado))
```

```
## Start:  AIC=234.3
## Species ~ 1
##
## trying + Sepal.Length
## # weights:  9 (4 variable)
```

```
## initial  value 115.354290
## iter  10 value 62.562717
## iter  20 value 62.200884
## iter  30 value 62.192424
## final  value 62.192209
## converged
## trying + Sepal.Width
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 93.617609
## final  value 93.616396
## converged
## trying + Petal.Length
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 14.092776
## iter  20 value 13.263608
## iter  30 value 13.152415
## iter  40 value 13.146931
## iter  50 value 13.146595
## iter  60 value 13.146399
## final  value 13.146285
## converged
## trying + Petal.Width
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 13.287178
## iter  20 value 12.062553
## iter  30 value 11.998988
## iter  40 value 11.953176
## iter  50 value 11.950654
## iter  60 value 11.947170
## iter  70 value 11.946489
## iter  80 value 11.944525
## iter  90 value 11.944244
## iter 100 value 11.943668
## final  value 11.943668
## stopped after 100 iterations
##                  Df       AIC
## + +Petal.Width    4  31.88734
## + +Petal.Length   4  34.29257
## + +Sepal.Length   4 132.38442
## + +Sepal.Width    4 195.23279
## <none>           2 234.30329
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 13.287178
## iter  20 value 12.062553
## iter  30 value 11.998988
## iter  40 value 11.953176
## iter  50 value 11.950654
## iter  60 value 11.947170
## iter  70 value 11.946489
## iter  80 value 11.944525
```

```
## iter  90 value 11.944244
## iter 100 value 11.943668
## final  value 11.943668
## stopped after 100 iterations
##
## Step:  AIC=31.89
## Species ~ Petal.Width
##
## trying + Sepal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 16.352816
## iter  20 value 12.454171
## iter  30 value 12.148159
## iter  40 value 11.905495
## iter  50 value 11.876547
## iter  60 value 11.875129
## iter  70 value 11.869701
## final  value 11.868493
## converged
## trying + Sepal.Width
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 11.080002
## iter  20 value 9.789672
## iter  30 value 9.730231
## iter  40 value 9.691618
## iter  50 value 9.679265
## iter  60 value 9.670540
## iter  70 value 9.668341
## iter  80 value 9.660126
## iter  90 value 9.658204
## iter 100 value 9.658102
## final  value 9.658102
## stopped after 100 iterations
## trying + Petal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 10.604971
## iter  20 value 8.547740
## iter  30 value 8.513403
## iter  40 value 8.500973
## iter  50 value 8.494082
## iter  60 value 8.492201
## iter  70 value 8.487866
## iter  80 value 8.486032
## iter  90 value 8.484822
## iter 100 value 8.483691
## final  value 8.483691
## stopped after 100 iterations
##                 Df      AIC
## + +Petal.Length  6 28.96738
## + +Sepal.Width   6 31.31620
## <none>           4 31.88734
```

```
## + +Sepal.Length  6 35.73699
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 10.604971
## iter  20 value 8.547740
## iter  30 value 8.513403
## iter  40 value 8.500973
## iter  50 value 8.494082
## iter  60 value 8.492201
## iter  70 value 8.487866
## iter  80 value 8.486032
## iter  90 value 8.484822
## iter 100 value 8.483691
## final  value 8.483691
## stopped after 100 iterations
##
## Step:  AIC=28.97
## Species ~ Petal.Width + Petal.Length
##
## trying + Sepal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
## trying + Sepal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 12.178204
## iter  20 value 5.050112
## iter  30 value 4.842539
## iter  40 value 4.731343
## iter  50 value 4.610612
## iter  60 value 4.597635
## final  value 4.596501
## converged
##                  Df      AIC
## + +Sepal.Length  8 24.79663
## + +Sepal.Width   8 25.19300
## <none>           6 28.96738
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
```

```
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
##
## Step:  AIC=24.8
## Species ~ Petal.Width + Petal.Length + Sepal.Length
##
## trying + Sepal.Width
## # weights:  18 (10 variable)
## initial  value 115.354290
## iter  10 value 14.037979
## iter  20 value 3.342288
## iter  30 value 2.503699
## iter  40 value 2.171547
## iter  50 value 2.099460
## iter  60 value 1.828506
## iter  70 value 0.904367
## iter  80 value 0.669147
## iter  90 value 0.622003
## iter 100 value 0.609416
## final  value 0.609416
## stopped after 100 iterations
##                Df      AIC
## + +Sepal.Width 10 21.21883
## <none>          8 24.79663
## # weights:  18 (10 variable)
## initial  value 115.354290
## iter  10 value 14.037979
## iter  20 value 3.342288
## iter  30 value 2.503699
## iter  40 value 2.171547
## iter  50 value 2.099460
## iter  60 value 1.828506
## iter  70 value 0.904367
## iter  80 value 0.669147
## iter  90 value 0.622003
## iter 100 value 0.609416
## final  value 0.609416
## stopped after 100 iterations
##
## Step:  AIC=21.22
## Species ~ Petal.Width + Petal.Length + Sepal.Length + Sepal.Width
```

```
modelo_stepwise <- step(modelo_nulo, direction = "both", scope =
 ↪  formula(modelo_ajustado))
```

```
## Start:  AIC=234.3
## Species ~ 1
```

```
##
## trying + Sepal.Length
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 62.562717
## iter  20 value 62.200884
## iter  30 value 62.192424
## final  value 62.192209
## converged
## trying + Sepal.Width
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 93.617609
## final  value 93.616396
## converged
## trying + Petal.Length
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 14.092776
## iter  20 value 13.263608
## iter  30 value 13.152415
## iter  40 value 13.146931
## iter  50 value 13.146595
## iter  60 value 13.146399
## final  value 13.146285
## converged
## trying + Petal.Width
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 13.287178
## iter  20 value 12.062553
## iter  30 value 11.998988
## iter  40 value 11.953176
## iter  50 value 11.950654
## iter  60 value 11.947170
## iter  70 value 11.946489
## iter  80 value 11.944525
## iter  90 value 11.944244
## iter 100 value 11.943668
## final  value 11.943668
## stopped after 100 iterations
##                 Df       AIC
## + +Petal.Width   4  31.88734
## + +Petal.Length  4  34.29257
## + +Sepal.Length  4 132.38442
## + +Sepal.Width   4 195.23279
## <none>           2 234.30329
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 13.287178
## iter  20 value 12.062553
## iter  30 value 11.998988
## iter  40 value 11.953176
## iter  50 value 11.950654
```

```
## iter  60 value 11.947170
## iter  70 value 11.946489
## iter  80 value 11.944525
## iter  90 value 11.944244
## iter 100 value 11.943668
## final  value 11.943668
## stopped after 100 iterations
##
## Step:  AIC=31.89
## Species ~ Petal.Width
##
## trying - Petal.Width
## # weights:  6 (2 variable)
## initial  value 115.354290
## final  value 115.151643
## converged
## trying + Sepal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 16.352816
## iter  20 value 12.454171
## iter  30 value 12.148159
## iter  40 value 11.905495
## iter  50 value 11.876547
## iter  60 value 11.875129
## iter  70 value 11.869701
## final  value 11.868493
## converged
## trying + Sepal.Width
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 11.080002
## iter  20 value 9.789672
## iter  30 value 9.730231
## iter  40 value 9.691618
## iter  50 value 9.679265
## iter  60 value 9.670540
## iter  70 value 9.668341
## iter  80 value 9.660126
## iter  90 value 9.658204
## iter 100 value 9.658102
## final  value 9.658102
## stopped after 100 iterations
## trying + Petal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 10.604971
## iter  20 value 8.547740
## iter  30 value 8.513403
## iter  40 value 8.500973
## iter  50 value 8.494082
## iter  60 value 8.492201
## iter  70 value 8.487866
## iter  80 value 8.486032
```

```
## iter  90 value 8.484822
## iter 100 value 8.483691
## final  value 8.483691
## stopped after 100 iterations
##                 Df       AIC
## + +Petal.Length  6  28.96738
## + +Sepal.Width   6  31.31620
## <none>           4  31.88734
## + +Sepal.Length  6  35.73699
## - Petal.Width    2 234.30329
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 10.604971
## iter  20 value 8.547740
## iter  30 value 8.513403
## iter  40 value 8.500973
## iter  50 value 8.494082
## iter  60 value 8.492201
## iter  70 value 8.487866
## iter  80 value 8.486032
## iter  90 value 8.484822
## iter 100 value 8.483691
## final  value 8.483691
## stopped after 100 iterations
##
## Step:  AIC=28.97
## Species ~ Petal.Width + Petal.Length
##
## trying - Petal.Width
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 14.092776
## iter  20 value 13.263608
## iter  30 value 13.152415
## iter  40 value 13.146931
## iter  50 value 13.146595
## iter  60 value 13.146399
## final  value 13.146285
## converged
## trying - Petal.Length
## # weights:  9 (4 variable)
## initial  value 115.354290
## iter  10 value 13.287178
## iter  20 value 12.062553
## iter  30 value 11.998988
## iter  40 value 11.953176
## iter  50 value 11.950654
## iter  60 value 11.947170
## iter  70 value 11.946489
## iter  80 value 11.944525
## iter  90 value 11.944244
## iter 100 value 11.943668
## final  value 11.943668
## stopped after 100 iterations
```

```
## trying + Sepal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
## trying + Sepal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 12.178204
## iter  20 value 5.050112
## iter  30 value 4.842539
## iter  40 value 4.731343
## iter  50 value 4.610612
## iter  60 value 4.597635
## final  value 4.596501
## converged
##                  Df      AIC
## + +Sepal.Length  8 24.79663
## + +Sepal.Width   8 25.19300
## <none>           6 28.96738
## - Petal.Length   4 31.88734
## - Petal.Width    4 34.29257
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
##
## Step:  AIC=24.8
## Species ~ Petal.Width + Petal.Length + Sepal.Length
##
## trying - Petal.Width
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 12.488824
```

```
## iter  20 value 6.973806
## iter  30 value 6.639299
## iter  40 value 6.563502
## iter  50 value 6.508959
## iter  60 value 6.488786
## iter  70 value 6.487951
## iter  80 value 6.486000
## iter  90 value 6.485869
## iter 100 value 6.485675
## final  value 6.485675
## stopped after 100 iterations
## trying - Petal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 16.352816
## iter  20 value 12.454171
## iter  30 value 12.148159
## iter  40 value 11.905495
## iter  50 value 11.876547
## iter  60 value 11.875129
## iter  70 value 11.869701
## final  value 11.868493
## converged
## trying - Sepal.Length
## # weights:  12 (6 variable)
## initial  value 115.354290
## iter  10 value 10.604971
## iter  20 value 8.547740
## iter  30 value 8.513403
## iter  40 value 8.500973
## iter  50 value 8.494082
## iter  60 value 8.492201
## iter  70 value 8.487866
## iter  80 value 8.486032
## iter  90 value 8.484822
## iter 100 value 8.483691
## final  value 8.483691
## stopped after 100 iterations
## trying + Sepal.Width
## # weights:  18 (10 variable)
## initial  value 115.354290
## iter  10 value 14.037979
## iter  20 value 3.342288
## iter  30 value 2.503699
## iter  40 value 2.171547
## iter  50 value 2.099460
## iter  60 value 1.828506
## iter  70 value 0.904367
## iter  80 value 0.669147
## iter  90 value 0.622003
## iter 100 value 0.609416
## final  value 0.609416
## stopped after 100 iterations
##                 Df      AIC
```

```
## + +Sepal.Width 10 21.21883
## <none>           8 24.79663
## - Petal.Width   6 24.97135
## - Sepal.Length  6 28.96738
## - Petal.Length  6 35.73699
## # weights:  18 (10 variable)
## initial  value 115.354290
## iter  10 value 14.037979
## iter  20 value 3.342288
## iter  30 value 2.503699
## iter  40 value 2.171547
## iter  50 value 2.099460
## iter  60 value 1.828506
## iter  70 value 0.904367
## iter  80 value 0.669147
## iter  90 value 0.622003
## iter 100 value 0.609416
## final  value 0.609416
## stopped after 100 iterations
##
## Step:  AIC=21.22
## Species ~ Petal.Width + Petal.Length + Sepal.Length + Sepal.Width
##
## trying - Petal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 17.073901
## iter  20 value 7.858714
## iter  30 value 6.707280
## iter  40 value 6.406272
## iter  50 value 6.381857
## iter  60 value 6.374101
## final  value 6.366160
## converged
## trying - Petal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 13.208650
## iter  20 value 9.268575
## iter  30 value 8.984624
## iter  40 value 8.955823
## iter  50 value 8.953394
## final  value 8.953389
## converged
## trying - Sepal.Length
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 12.178204
## iter  20 value 5.050112
## iter  30 value 4.842539
## iter  40 value 4.731343
## iter  50 value 4.610612
## iter  60 value 4.597635
## final  value 4.596501
```

```
## converged
## trying - Sepal.Width
## # weights:  15 (8 variable)
## initial  value 115.354290
## iter  10 value 9.870302
## iter  20 value 4.460007
## iter  30 value 4.407318
## iter  40 value 4.406430
## iter  50 value 4.403603
## iter  60 value 4.400206
## iter  70 value 4.398808
## iter  80 value 4.398415
## iter  90 value 4.398384
## iter 100 value 4.398316
## final  value 4.398316
## stopped after 100 iterations
##                Df      AIC
## <none>         10 21.21883
## - Sepal.Width   8 24.79663
## - Sepal.Length  8 25.19300
## - Petal.Width   8 28.73232
## - Petal.Length  8 33.90678
```

5) Con el modelo resultante del apartado anterior, obtener medidas de bondad del ajuste e indicar si el modelo es significativo.

6) Veamos ahora el problema de Regresión Multinomial como un problema de clasificación. Obtener la clase predicha para los datos del conjunto de prueba.

7) Obtener la matriz de confusión para el conjunto de prueba y medir la eficiencia del clasificador con el "accuracy" (número de aciertos en la clasificación dividido entre número de datos totales).