

Análisis Discriminante

Lino Oswaldo Sanchez

21/5/2022

Introducción

El análisis discriminante tiene como objetivo separar (discriminar) un grupo de 2 o más poblaciones hallando diferencias entre ellas con una función discriminante. Este se puede llamar análisis discriminante descriptivo; existe el análisis predictivo donde la Predicción o asignación de un objeto en uno de entre varios grupos previamente definidos, ubicando un objeto o unidad muestral en uno o varios grupos de acuerdo a las reglas de clasificación.

Es un criterio que permite asignar las observaciones de las variables sobre un individuo W y que a menudo es planteado mediante una función discriminante $D(x_1, \dots, x_p)$.

La regla de clasificación es:

Sí $D(x_1, \dots, x_p) > 0$ asignamos W a ω_1 ,

en caso contrario, asignamos W a ω_2 .

Base de datos

librerías necesarias

```
library(MASS)
```

```
Z<-as.data.frame(iris)
head(Z)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	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	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa

Traemos la base y visualizamos los nombres de las variables para tener una idea de que variables están dentro, encontramos datos sobre los tipos de flores tales como largo del petalo, ancho del petalo, largo del sepalo, ancho del sepalo y especie.

Tenemos 5 variables de las cuales 4 son cuantitativas numéricas y una carácter.

- Definimos la matriz de datos y la variable respuesta con las categorías.

De la uno a la 4 las cuantitativas

```
x<-Z[,1:4]
x
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
## 1	5.1	3.5	1.4	0.2
## 2	4.9	3.0	1.4	0.2
## 3	4.7	3.2	1.3	0.2
## 4	4.6	3.1	1.5	0.2
## 5	5.0	3.6	1.4	0.2
## 6	5.4	3.9	1.7	0.4
## 7	4.6	3.4	1.4	0.3
## 8	5.0	3.4	1.5	0.2
## 9	4.4	2.9	1.4	0.2
## 10	4.9	3.1	1.5	0.1
## 11	5.4	3.7	1.5	0.2
## 12	4.8	3.4	1.6	0.2
## 13	4.8	3.0	1.4	0.1
## 14	4.3	3.0	1.1	0.1
## 15	5.8	4.0	1.2	0.2
## 16	5.7	4.4	1.5	0.4
## 17	5.4	3.9	1.3	0.4
## 18	5.1	3.5	1.4	0.3
## 19	5.7	3.8	1.7	0.3
## 20	5.1	3.8	1.5	0.3
## 21	5.4	3.4	1.7	0.2
## 22	5.1	3.7	1.5	0.4
## 23	4.6	3.6	1.0	0.2
## 24	5.1	3.3	1.7	0.5
## 25	4.8	3.4	1.9	0.2
## 26	5.0	3.0	1.6	0.2
## 27	5.0	3.4	1.6	0.4
## 28	5.2	3.5	1.5	0.2
## 29	5.2	3.4	1.4	0.2
## 30	4.7	3.2	1.6	0.2
## 31	4.8	3.1	1.6	0.2
## 32	5.4	3.4	1.5	0.4
## 33	5.2	4.1	1.5	0.1
## 34	5.5	4.2	1.4	0.2
## 35	4.9	3.1	1.5	0.2
## 36	5.0	3.2	1.2	0.2
## 37	5.5	3.5	1.3	0.2
## 38	4.9	3.6	1.4	0.1
## 39	4.4	3.0	1.3	0.2
## 40	5.1	3.4	1.5	0.2
## 41	5.0	3.5	1.3	0.3
## 42	4.5	2.3	1.3	0.3
## 43	4.4	3.2	1.3	0.2
## 44	5.0	3.5	1.6	0.6
## 45	5.1	3.8	1.9	0.4
## 46	4.8	3.0	1.4	0.3
## 47	5.1	3.8	1.6	0.2
## 48	4.6	3.2	1.4	0.2

## 49	5.3	3.7	1.5	0.2
## 50	5.0	3.3	1.4	0.2
## 51	7.0	3.2	4.7	1.4
## 52	6.4	3.2	4.5	1.5
## 53	6.9	3.1	4.9	1.5
## 54	5.5	2.3	4.0	1.3
## 55	6.5	2.8	4.6	1.5
## 56	5.7	2.8	4.5	1.3
## 57	6.3	3.3	4.7	1.6
## 58	4.9	2.4	3.3	1.0
## 59	6.6	2.9	4.6	1.3
## 60	5.2	2.7	3.9	1.4
## 61	5.0	2.0	3.5	1.0
## 62	5.9	3.0	4.2	1.5
## 63	6.0	2.2	4.0	1.0
## 64	6.1	2.9	4.7	1.4
## 65	5.6	2.9	3.6	1.3
## 66	6.7	3.1	4.4	1.4
## 67	5.6	3.0	4.5	1.5
## 68	5.8	2.7	4.1	1.0
## 69	6.2	2.2	4.5	1.5
## 70	5.6	2.5	3.9	1.1
## 71	5.9	3.2	4.8	1.8
## 72	6.1	2.8	4.0	1.3
## 73	6.3	2.5	4.9	1.5
## 74	6.1	2.8	4.7	1.2
## 75	6.4	2.9	4.3	1.3
## 76	6.6	3.0	4.4	1.4
## 77	6.8	2.8	4.8	1.4
## 78	6.7	3.0	5.0	1.7
## 79	6.0	2.9	4.5	1.5
## 80	5.7	2.6	3.5	1.0
## 81	5.5	2.4	3.8	1.1
## 82	5.5	2.4	3.7	1.0
## 83	5.8	2.7	3.9	1.2
## 84	6.0	2.7	5.1	1.6
## 85	5.4	3.0	4.5	1.5
## 86	6.0	3.4	4.5	1.6
## 87	6.7	3.1	4.7	1.5
## 88	6.3	2.3	4.4	1.3
## 89	5.6	3.0	4.1	1.3
## 90	5.5	2.5	4.0	1.3
## 91	5.5	2.6	4.4	1.2
## 92	6.1	3.0	4.6	1.4
## 93	5.8	2.6	4.0	1.2
## 94	5.0	2.3	3.3	1.0
## 95	5.6	2.7	4.2	1.3
## 96	5.7	3.0	4.2	1.2
## 97	5.7	2.9	4.2	1.3
## 98	6.2	2.9	4.3	1.3
## 99	5.1	2.5	3.0	1.1
## 100	5.7	2.8	4.1	1.3
## 101	6.3	3.3	6.0	2.5
## 102	5.8	2.7	5.1	1.9

## 103	7.1	3.0	5.9	2.1
## 104	6.3	2.9	5.6	1.8
## 105	6.5	3.0	5.8	2.2
## 106	7.6	3.0	6.6	2.1
## 107	4.9	2.5	4.5	1.7
## 108	7.3	2.9	6.3	1.8
## 109	6.7	2.5	5.8	1.8
## 110	7.2	3.6	6.1	2.5
## 111	6.5	3.2	5.1	2.0
## 112	6.4	2.7	5.3	1.9
## 113	6.8	3.0	5.5	2.1
## 114	5.7	2.5	5.0	2.0
## 115	5.8	2.8	5.1	2.4
## 116	6.4	3.2	5.3	2.3
## 117	6.5	3.0	5.5	1.8
## 118	7.7	3.8	6.7	2.2
## 119	7.7	2.6	6.9	2.3
## 120	6.0	2.2	5.0	1.5
## 121	6.9	3.2	5.7	2.3
## 122	5.6	2.8	4.9	2.0
## 123	7.7	2.8	6.7	2.0
## 124	6.3	2.7	4.9	1.8
## 125	6.7	3.3	5.7	2.1
## 126	7.2	3.2	6.0	1.8
## 127	6.2	2.8	4.8	1.8
## 128	6.1	3.0	4.9	1.8
## 129	6.4	2.8	5.6	2.1
## 130	7.2	3.0	5.8	1.6
## 131	7.4	2.8	6.1	1.9
## 132	7.9	3.8	6.4	2.0
## 133	6.4	2.8	5.6	2.2
## 134	6.3	2.8	5.1	1.5
## 135	6.1	2.6	5.6	1.4
## 136	7.7	3.0	6.1	2.3
## 137	6.3	3.4	5.6	2.4
## 138	6.4	3.1	5.5	1.8
## 139	6.0	3.0	4.8	1.8
## 140	6.9	3.1	5.4	2.1
## 141	6.7	3.1	5.6	2.4
## 142	6.9	3.1	5.1	2.3
## 143	5.8	2.7	5.1	1.9
## 144	6.8	3.2	5.9	2.3
## 145	6.7	3.3	5.7	2.5
## 146	6.7	3.0	5.2	2.3
## 147	6.3	2.5	5.0	1.9
## 148	6.5	3.0	5.2	2.0
## 149	6.2	3.4	5.4	2.3
## 150	5.9	3.0	5.1	1.8

La 5 que es la categórica que contiene la especie de la flor.

```
y<-Z[,5]
```

- Definir como n y p el número de flores y variables

```
n<-nrow(x)
p<-ncol(x)
```

Aplicación del Análisis discriminante lineal (LDA)

Con Cross validation(validación cruzada) cv: clasificación optima (TRUE).

```
lda.iris<-lda(Z$Species~.,data=Z,CV=TRUE)
```

Y en función de todas las observaciones con clasificación optima y clasificación cruzada en este objeto se encontrarán las clase y las probabilidades posteriori.

- *lda.iris\$class* contiene las clasificaciones hechas por CV usando LDA.

```
lda.iris$class
```

```
## [1] setosa setosa setosa setosa setosa setosa
## [7] setosa setosa setosa setosa setosa setosa
## [13] setosa setosa setosa setosa setosa setosa
## [19] setosa setosa setosa setosa setosa setosa
## [25] setosa setosa setosa setosa setosa setosa
## [31] setosa setosa setosa setosa setosa setosa
## [37] setosa setosa setosa setosa setosa setosa
## [43] setosa setosa setosa setosa setosa setosa
## [49] setosa setosa versicolor versicolor versicolor versicolor
## [55] versicolor versicolor versicolor versicolor versicolor versicolor
## [61] versicolor versicolor versicolor versicolor versicolor versicolor
## [67] versicolor versicolor versicolor versicolor virginica versicolor
## [73] versicolor versicolor versicolor versicolor versicolor versicolor
## [79] versicolor versicolor versicolor versicolor versicolor virginica
## [85] versicolor versicolor versicolor versicolor versicolor versicolor
## [91] versicolor versicolor versicolor versicolor versicolor versicolor
## [97] versicolor versicolor versicolor versicolor virginica virginica
## [103] virginica virginica virginica virginica virginica virginica
## [109] virginica virginica virginica virginica virginica virginica
## [115] virginica virginica virginica virginica virginica virginica
## [121] virginica virginica virginica virginica virginica virginica
## [127] virginica virginica virginica virginica virginica virginica
## [133] virginica versicolor virginica virginica virginica virginica
## [139] virginica virginica virginica virginica virginica virginica
## [145] virginica virginica virginica virginica virginica virginica
## Levels: setosa versicolor virginica
```

vemos las clase hechas por la clasificación cruzada

Creación de la tabla de clasificaciones buenas y malas

```
table.lda<-table(y,lda.iris$class)
table.lda
```

```
##
## y          setosa versicolor virginica
##  setosa      50         0         0
##  versicolor  0         48         2
##  virginica   0         1        49
```

Al interpretar la tabla vemos que para la especie setosa la clasifica correctamente al 100 % mientras que para versicolor y 48 son clasificadas como tal y 2 de esas misma especie las clasifica cómo virginica, para la especie virginica 49 son clasificadas como tal y una es clasificada como versicolor.

El erro de clasificación es aparentemente pequeño.

- Proporción de errores

```
mis.lda<- n-sum(y==lda.iris$class)
mis.lda/n
```

```
## [1] 0.02
```

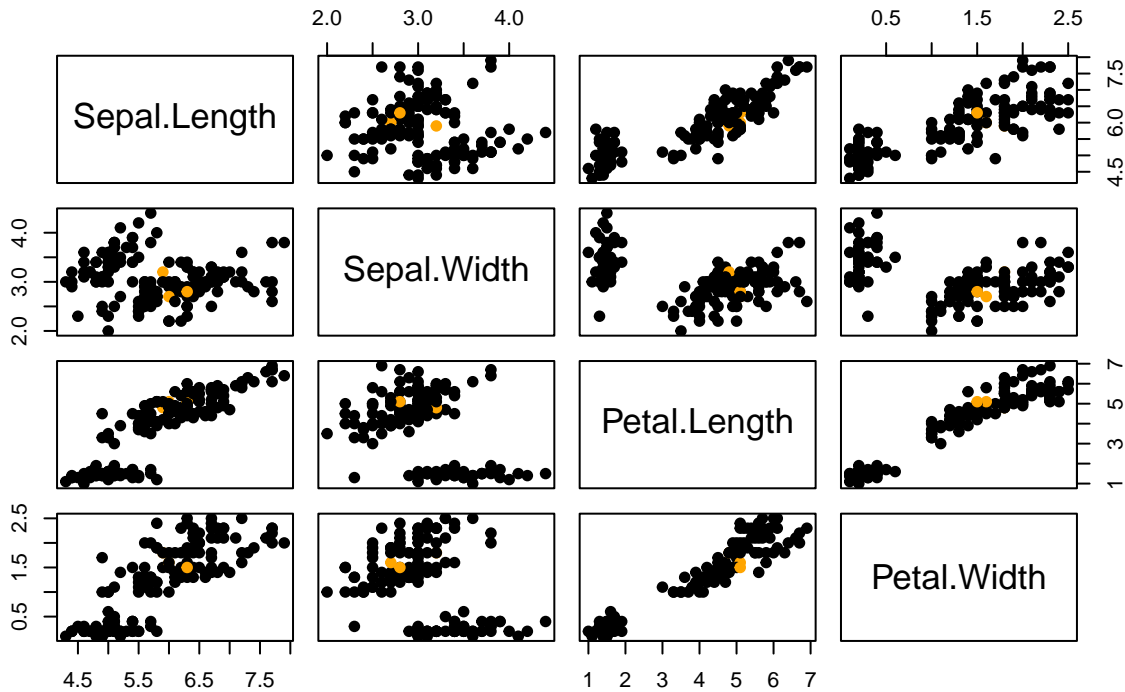
El error es de: 0.02.Se obtiene calculando que es n - la suma (la probabilidad de los errores dentro de `lda.iris$class`) y el resultado se divide entre “n”

scatter plot para las clacificaciones

Buenas clasificaciones en negro y malas en rojo

```
col.lda.iris<-c("orange","black")[1*(y==lda.iris$class)+1]
pairs(x,main="Buena clasificación (negro), mala clasificación (naranja)",
      pch=19,col=col.lda.iris)
```

Buena clasificación (negro), mala clasificación (naranja)



El gráfico nos deja ver que en su mayoría las observaciones están bien clasificadas

Probabilidad de pertenencia a uno de los tres grupos

En esta lista podemos identificar los que no están clasificados bien que por pura inferencia podrían ser catalogados como atípicos, o que comparten características con la especie a la que se le confunde.

```
##          setosa  versicolor  virginica
## 1  1.000000e+00 5.087494e-22 4.385241e-42
## 2  1.000000e+00 9.588256e-18 8.888069e-37
## 3  1.000000e+00 1.983745e-19 8.606982e-39
## 4  1.000000e+00 1.505573e-16 5.101765e-35
## 5  1.000000e+00 2.075670e-22 1.739832e-42
## 6  1.000000e+00 5.332271e-21 8.674906e-40
## 7  1.000000e+00 1.498839e-18 3.999205e-37
## 8  1.000000e+00 5.268133e-20 1.983027e-39
## 9  1.000000e+00 2.280729e-15 1.293376e-33
## 10 1.000000e+00 1.504085e-18 5.037348e-38
## 11 1.000000e+00 1.296140e-23 4.023338e-44
## 12 1.000000e+00 2.171874e-18 3.223111e-37
## 13 1.000000e+00 1.996136e-18 6.109118e-38
## 14 1.000000e+00 1.604055e-19 2.549802e-39
## 15 1.000000e+00 2.843397e-31 1.593594e-54
## 16 1.000000e+00 2.330545e-28 3.074132e-49
## 17 1.000000e+00 5.136116e-25 3.269819e-45
## 18 1.000000e+00 5.747697e-21 2.253825e-40
```

```

## 19 1.000000e+00 2.187125e-22 4.069438e-42
## 20 1.000000e+00 3.297882e-22 9.802494e-42
## 21 1.000000e+00 1.757286e-19 8.150916e-39
## 22 1.000000e+00 2.027767e-20 3.730752e-39
## 23 1.000000e+00 5.650696e-25 6.509776e-46
## 24 1.000000e+00 8.618517e-15 7.014744e-32
## 25 1.000000e+00 1.520334e-15 1.857885e-33
## 26 1.000000e+00 2.936141e-16 8.159510e-35
## 27 1.000000e+00 4.557392e-17 5.510803e-35
## 28 1.000000e+00 2.079675e-21 2.831513e-41
## 29 1.000000e+00 1.232321e-21 1.082692e-41
## 30 1.000000e+00 1.153050e-16 4.267126e-35
## 31 1.000000e+00 2.584595e-16 9.537258e-35
## 32 1.000000e+00 2.878754e-19 5.473623e-38
## 33 1.000000e+00 2.247070e-27 4.047137e-49
## 34 1.000000e+00 2.620949e-29 1.970538e-51
## 35 1.000000e+00 1.493279e-17 2.047516e-36
## 36 1.000000e+00 2.146308e-21 1.550216e-41
## 37 1.000000e+00 1.673983e-24 1.322398e-45
## 38 1.000000e+00 3.810942e-23 9.131835e-44
## 39 1.000000e+00 5.423320e-17 1.146137e-35
## 40 1.000000e+00 2.414191e-20 6.552342e-40
## 41 1.000000e+00 1.417602e-21 3.569675e-41
## 42 1.000000e+00 8.956712e-11 4.968454e-28
## 43 1.000000e+00 2.125837e-18 2.395462e-37
## 44 1.000000e+00 1.101293e-15 1.403899e-32
## 45 1.000000e+00 2.285363e-17 5.214629e-35
## 46 1.000000e+00 2.087086e-16 1.027948e-34
## 47 1.000000e+00 2.588201e-22 3.634491e-42
## 48 1.000000e+00 3.643000e-18 4.504970e-37
## 49 1.000000e+00 3.000767e-23 1.346233e-43
## 50 1.000000e+00 3.171862e-20 7.860312e-40
## 51 3.157725e-18 9.998716e-01 1.284247e-04
## 52 1.753919e-19 9.991816e-01 8.184018e-04
## 53 2.551962e-22 9.951044e-01 4.895626e-03
## 54 2.742687e-22 9.995996e-01 4.004477e-04
## 55 4.854978e-23 9.951404e-01 4.859638e-03
## 56 9.575747e-23 9.982973e-01 1.702702e-03
## 57 4.467689e-22 9.838631e-01 1.613691e-02
## 58 5.922943e-14 9.999999e-01 8.584221e-08
## 59 8.088509e-20 9.998655e-01 1.344590e-04
## 60 1.767441e-20 9.994314e-01 5.686054e-04
## 61 3.330661e-18 9.999987e-01 1.314516e-06
## 62 8.331100e-20 9.991631e-01 8.369389e-04
## 63 4.614428e-18 9.999989e-01 1.117671e-06
## 64 1.290071e-23 9.939163e-01 6.083745e-03
## 65 5.229707e-14 9.999984e-01 1.593028e-06
## 66 3.393529e-17 9.999528e-01 4.721492e-05
## 67 7.983370e-24 9.763990e-01 2.360097e-02
## 68 3.119288e-16 9.999991e-01 8.659241e-07
## 69 3.847473e-28 9.390462e-01 6.095377e-02
## 70 1.678698e-17 9.999966e-01 3.360127e-06
## 71 1.302246e-28 1.772727e-01 8.227273e-01
## 72 1.113263e-16 9.999902e-01 9.801197e-06

```



```

## 73 1.634947e-29 7.868347e-01 2.131653e-01
## 74 3.331093e-22 9.995073e-01 4.926830e-04
## 75 1.013127e-17 9.999741e-01 2.594176e-05
## 76 2.949236e-18 9.999081e-01 9.193549e-05
## 77 7.224891e-23 9.979459e-01 2.054146e-03
## 78 2.386376e-27 6.569495e-01 3.430505e-01
## 79 4.473658e-23 9.922840e-01 7.716012e-03
## 80 7.145460e-12 1.000000e+00 1.241414e-08
## 81 1.333306e-17 9.999970e-01 3.044209e-06
## 82 1.119894e-15 9.999997e-01 2.916503e-07
## 83 1.748156e-16 9.999961e-01 3.876682e-06
## 84 1.125494e-33 9.924153e-02 9.007585e-01
## 85 1.191672e-24 9.474667e-01 5.253333e-02
## 86 1.983291e-20 9.924721e-01 7.527887e-03
## 87 4.531906e-21 9.980100e-01 1.989996e-03
## 88 2.035626e-23 9.993358e-01 6.642410e-04
## 89 7.813451e-18 9.999440e-01 5.603286e-05
## 90 8.212308e-21 9.998033e-01 1.967487e-04
## 91 6.631189e-23 9.992802e-01 7.197827e-04
## 92 7.049062e-22 9.979525e-01 2.047473e-03
## 93 4.490728e-18 9.999881e-01 1.188058e-05
## 94 2.600275e-14 9.999999e-01 8.745690e-08
## 95 6.422939e-21 9.996751e-01 3.248823e-04
## 96 2.159263e-17 9.999804e-01 1.956029e-05
## 97 3.823305e-19 9.998801e-01 1.199041e-04
## 98 2.089502e-18 9.999504e-01 4.963639e-05
## 99 9.013113e-11 1.000000e+00 9.943306e-09
## 100 6.167377e-19 9.999219e-01 7.813051e-05
## 101 1.335977e-53 3.188548e-09 1.000000e+00
## 102 9.949508e-38 1.209398e-03 9.987906e-01
## 103 1.950796e-42 2.774428e-05 9.999723e-01
## 104 3.081602e-38 1.232592e-03 9.987674e-01
## 105 5.411117e-46 1.807449e-06 9.999982e-01
## 106 5.887455e-50 5.662591e-07 9.999994e-01
## 107 1.203272e-32 8.794800e-02 9.120520e-01
## 108 1.774038e-42 1.735541e-04 9.998264e-01
## 109 1.924345e-42 2.617818e-04 9.997382e-01
## 110 1.851248e-46 1.352651e-07 9.999999e-01
## 111 4.379051e-32 1.446014e-02 9.855399e-01
## 112 2.052671e-37 1.776421e-03 9.982236e-01
## 113 9.704392e-39 2.172029e-04 9.997828e-01
## 114 2.386650e-40 2.251253e-04 9.997749e-01
## 115 8.048237e-46 8.410965e-07 9.999992e-01
## 116 1.008588e-39 2.840103e-05 9.999716e-01
## 117 2.811294e-35 6.595206e-03 9.934048e-01
## 118 7.282186e-45 1.296566e-06 9.999987e-01
## 119 1.004644e-64 2.647509e-10 1.000000e+00
## 120 3.160887e-33 3.033047e-01 6.966953e-01
## 121 1.719583e-42 6.688965e-06 9.999933e-01
## 122 6.252717e-37 9.870164e-04 9.990130e-01
## 123 2.627103e-51 7.704580e-07 9.999992e-01
## 124 1.504499e-31 1.070121e-01 8.929879e-01
## 125 3.688147e-39 9.571422e-05 9.999043e-01
## 126 2.426533e-36 3.398007e-03 9.966020e-01

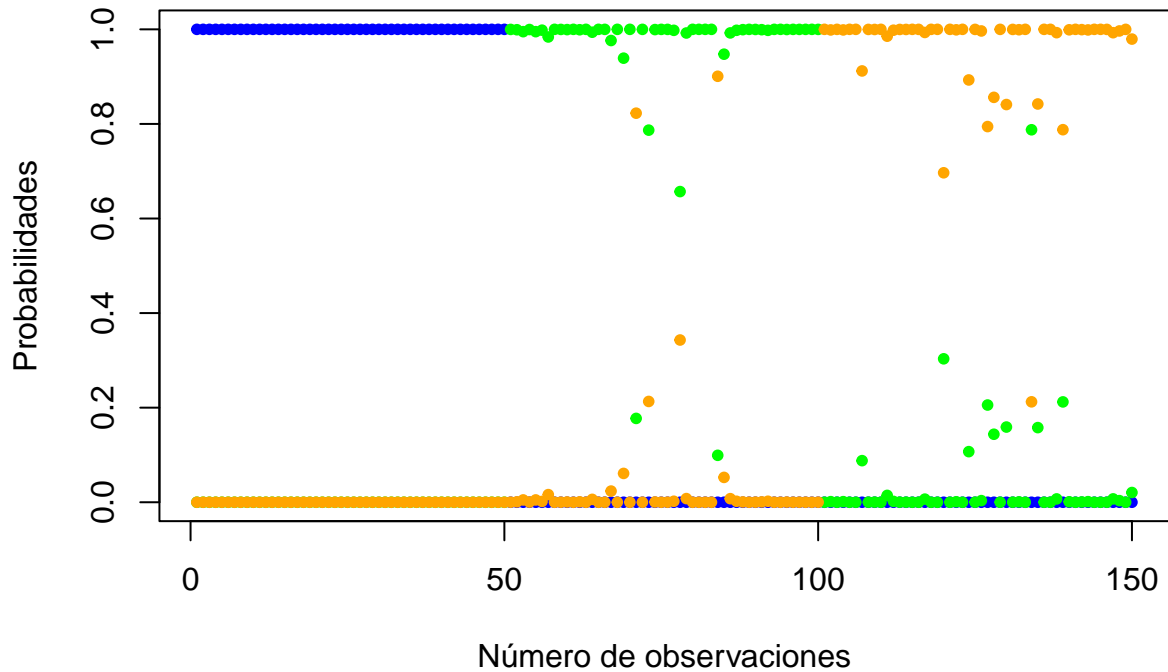
```

```
## 127 3.865436e-30 2.055755e-01 7.944245e-01
## 128 3.606381e-30 1.437670e-01 8.562330e-01
## 129 8.371636e-44 1.376281e-05 9.999862e-01
## 130 2.937738e-32 1.589920e-01 8.410080e-01
## 131 6.294581e-42 1.714027e-04 9.998286e-01
## 132 5.466934e-36 7.736441e-04 9.992264e-01
## 133 1.208158e-45 3.051435e-06 9.999969e-01
## 134 5.464475e-29 7.876238e-01 2.123762e-01
## 135 9.884011e-35 1.578198e-01 8.421802e-01
## 136 6.515088e-46 1.990735e-06 9.999980e-01
## 137 2.840394e-44 7.895048e-07 9.999992e-01
## 138 7.160822e-35 7.053731e-03 9.929463e-01
## 139 1.782247e-29 2.122042e-01 7.877958e-01
## 140 3.640914e-36 9.289807e-04 9.990710e-01
## 141 5.881132e-45 1.108009e-06 9.999989e-01
## 142 2.122304e-35 6.157433e-04 9.993843e-01
## 143 9.949508e-38 1.209398e-03 9.987906e-01
## 144 9.585800e-46 9.978596e-07 9.999990e-01
## 145 2.206003e-46 2.038879e-07 9.999998e-01
## 146 1.133074e-38 8.851900e-05 9.999115e-01
## 147 8.781586e-36 7.084468e-03 9.929155e-01
## 148 7.108984e-35 3.342993e-03 9.966570e-01
## 149 3.096565e-40 1.338572e-05 9.999866e-01
## 150 3.585667e-33 2.058806e-02 9.794119e-01
```

Gráfico de probabilidades

```
plot(1:n, lda.iris$posterior[,1],
     main="Probabilidades a posteriori",
     pch=20, col="blue",
     xlab="Número de observaciones", ylab="Probabilidades")
points(1:n, lda.iris$posterior[,2],
       pch=20, col="green")
points(1:n, lda.iris$posterior[,3],
       pch=20, col="orange")
```

Probabilidades a posteriori



Lo que observamos en el gráfico es la probabilidad de clasificación para las especies y observamos en azul a setosa que esta bien clasificada con una probabilidad de 1, para versicolor en verde observamos que no esta tan bien clasificada pues la probabilidad no es de uno para todas las observaciones y para virginica naranja, de igual forma no todas las observaciones están bien clasificadas o su probabilidad de clasificación no es igual a 1.

La parte de baja del gráfico nos muestra la probabilidad de que una especie este clasificada en otra y como vemos para setosa es de cero, pero no así para versicolor y virginica vemos probabilidades ciertamente un poco altas; un tip para clasificar es que toda observación que esta por debajo de 0.5 no será clasificado y asignar la observación a un grupo aunque no sea de la especie pues si comparte una alta probabilidad de pertenecer es mejor dejarlo allí que moverlo al grupo de pertenencia que debería ser.