

Práctica 2

Probabilidad y variable aleatoria



```
datos  
m <- naiveBayes(NOTA_FINAL.JUNIO...0.10, ~., data = datos)  
predictions <- predict(m, datos)  
table(predictions, datos$NOTA_FINAL.JUNIO...0.10.)  
table(datos$NOTA_FINAL.JUNIO...0.10.)  
  
#Ejercicio 2  
datos$NOTA_FINAL.JUNIO...0.10.<-cut(datos$NOTA_FINAL.JUNIO...0.10., breaks = c(0,5,10))  
m<-naiveBayes(datos$NOTA_FINAL.JUNIO...0.10.-datos$Grupo+datos$Práctica.1..0.2.3.,data=datos[,c(1:9,10)])  
  
#Ejercicio 3  
  
datos <- read.csv("Notas1.csv", sep=";", dec=",", header=T)  
  
for(i in c(2:4)){(datos[,i])*3*5}  
datos$Pr?.ctica.1..0.2.3. <- datos$Pr?.ctica.1..0.2.3.*(10/(2/3))  
datos$Cuestionario.temas.1.y.2..0.10.<-datos$Cuestionario.temas.1.y.2..0.10.  
datos$Cuestionario.tipo.test.temas.3.y.4...0.10.<-datos$Cuestionario.tipo.test.temas.3.y.4...0.10.  
  
datos$Cuestionario.tipo.test.temas.5.y.6...0.10.<-NULL  
datos$Pr?.ctica.2...0.2.3.<-NULL  
datos$Pr?.ctica.3...0.2.3.<-NULL  
datos$Total.cuestionarios..0.1.<-NULL  
datos$EXAMEN.JUNIO..0.7.<-NULL  
datos$NOTA_FINAL.SEPTIEMBRE..0.10. <- NULL  
datos$EXAMEN.SEPTIEMBRE..0.7.<-NULL  
datos$EXAMEN.JUNIO..0.7.<-NULL
```

Grupo B

Lisardo Carretero Colmenar
Luis Díaz González
David Gómez López
María Trinidad Gómez Hernández
Fátima Paterna Roda

Índice

Objetivo 1	2
Ilustración 1 cambio de variables cuantitativas a cualitativas.....	2
Ilustración 2 probabilidades condicionadas.....	2
Ilustración 3 Naive Bayes	4
Objetivo 2	5
Ilustración 4 cambio de NA por "no presentado"	5
Ilustración 5 cambio de variables cuantitativas a cualitativas.....	5
Ilustración 6 Naive Bayes	5
Ilustración 7 predicción total cuestionarios	7
Ilustración 8 predicción total prácticas	8
Ilustración 9 paso de suspenso/aprobado a suspenso, aprobado, notable y sobresaliente	8
Objetivo 3	8
Ilustración 10 tabla de predicciones	9
Ilustración 11 proceso de los calculos.....	10
Ilustración 12 predicciones de los datos de los componentes	10

Objetivo 1

1. Mediante el siguiente código pasamos las variables cuantitativas a cualitativas con los valores suspento y aprobado.

```
for(i in c(2:11)){datos[is.na(datos[,i]),i]<-0}
for(i in c(2:11)){
  datos[,i] <- cut(datos[,i], breaks = c(0, 5, 10), labels = c("Suspento", "Aprobado"))}
```

Ilustración 1 cambio de variables cuantitativas a cualitativas

2. Para calcular las probabilidades condicionadas a lo que se pide en el ejercicio 2 del objetivo 1 utilizamos:

```
m <- naiveBayes(NOTA.FINAL.JUNIO...0.10. ~. , data = datos)
m
predictions <- predict(m, datos)
table(predictions, datos$NOTA.FINAL.JUNIO...0.10.)
table(datos$NOTA.FINAL.JUNIO...0.10.)
```

Ilustración 2 probabilidades condicionadas

A priori las probabilidades de suspender y aprobar es del 71,8% y 28,2% respectivamente.

- GRUPOS

La probabilidad de aprobar siendo del grupo A es del 38%

La probabilidad de aprobar siendo del grupo B es del 61,9%

La probabilidad de suspender siendo del grupo A es del 54,2%

La probabilidad de suspender siendo del grupo B es del 45,8%

- PRACTICA 1

La probabilidad de aprobar habiendo aprobado es del 82,5%

La probabilidad de aprobar habiendo suspendido es del 17,5%

La probabilidad de suspender habiendo suspendido es del 23,8%

La probabilidad de suspender habiendo aprobado es del 76,2%

- PRACTICA 2

La probabilidad de aprobar habiendo aprobado es del 86,5%

La probabilidad de aprobar habiendo suspendido es del 13,5%

La probabilidad de suspender habiendo suspendido es del 43,9%

La probabilidad de suspender habiendo aprobado es del 56,1%

- PRACTICA 3

La probabilidad de aprobar habiendo aprobado es del 87,8%
La probabilidad de aprobar habiendo suspendido es del 12,2%
La probabilidad de suspender habiendo suspendido es del 35,6%
La probabilidad de suspender habiendo aprobado es del 64,4%

- TOTAL PRACTICAS

La probabilidad de aprobar habiendo aprobado es del 78,6%
La probabilidad de aprobar habiendo suspendido es del 21,4%
La probabilidad de suspender habiendo suspendido es del 56,4%
La probabilidad de suspender habiendo aprobado es del 43,6%

- CUESTIONARIO 1

La probabilidad de aprobar habiendo aprobado es del 61,2%
La probabilidad de aprobar habiendo suspendido es del 38,5%
La probabilidad de suspender habiendo suspendido es del 66,7%
La probabilidad de suspender habiendo aprobado es del 36,7%

- CUESTIONARIO 2

La probabilidad de aprobar habiendo aprobado es del 34,3%
La probabilidad de aprobar habiendo suspendido es del 65,7%
La probabilidad de suspender habiendo suspendido es del 63,3%
La probabilidad de suspender habiendo aprobado es del 36,7%

- CUESTIONARIO 3

La probabilidad de aprobar habiendo aprobado es del 44,8%
La probabilidad de aprobar habiendo suspendido es del 69%
La probabilidad de suspender habiendo suspendido es del 69,2%
La probabilidad de suspender habiendo aprobado es del 30,8%

- TOTAL CUESTIONARIO

La probabilidad de aprobar habiendo aprobado es del 31%
La probabilidad de aprobar habiendo suspendido es del 69%
La probabilidad de suspender habiendo suspendido es del 90%
La probabilidad de suspender habiendo aprobado es del 10%

- EXAMEN JUNIO

La probabilidad de aprobar habiendo aprobado es del 76,2%

La probabilidad de aprobar habiendo suspendido es del 23,8%

La probabilidad de suspender habiendo suspendido es del 97,1%

La probabilidad de suspender habiendo aprobado es del 2,9%

Con todo estos datos obtenemos una predicción de que en junio habrá unas 107 suspensos y 42 aprobados.

La predicción cuenta con un margen de error, puesto que con naiveBayes obtenemos un resultado de:

El número de alumnos que aprueben habiendo aprobado es de 32, una probabilidad de 21,48%

El número de alumnos que aprueben habiendo suspendido es de 4, una probabilidad de 2,68%

El número de alumnos que suspendan habiendo suspendido es de 103, una probabilidad de 69,13%

El número de alumnos que suspendan habiendo aprobado es de 10, una probabilidad de 6,71%

3. Evaluamos el cambio de las medidas anteriores utilizando el clasificador de Naive Bayes con los siguientes grupos de variables:

```
n <- naiveBayes(NOTA.FINAL.JUNIO...0.10. ~ datos$?.Grupo, data = datos)
n
predictions <- predict(n, datos)
table(predictions, datos$NOTA.FINAL.JUNIO...0.10.)
```

Ilustración 3 Naive Bayes

- Practica 1 y cuestionario 1

	Suspensos	Aprobados
Suspensos	100	33
Aprobados	7	9

- Practica 1, practica 2, cuestionario 1 y cuestionario 2

	Suspensos	Aprobados
Suspensos	101	27
Aprobados	6	15

- c. Utilizando toda la información de grupo, practicas y cuestionarios

	Suspenso	Aprobado
Suspenso	104	31
Aprobado	3	11

Con esto comprobamos que cuantas mas variables cogemos para realizar la predicción, más se ajusta al resultado obtenido en el anterior objetivo, lo que significa que es predicción mas acertada.

Objetivo 2

1. Se considera que algún alumno pueda aparecer como no presentado, por lo que cambiamos los NA a “no presentado”.
2. Transformamos las variables cuantitativas a cualitativas con los valores suspenso, aprobado, notable y sobresaliente.

Ilustración 4 cambio de NA por "no presentado"

```
for(i in c(2:11)){datos[is.na(datos[,i]),i]<- -1}
for(i in c(2:11)){datos[,i] <- cut(datos[,i],breaks = c(-1, 0, 5, 10), labels = c("No presentado","Suspenso", "Aprobado"))}
datos$TOTAL.pr?.cticas..0.2.
```

```
for(i in c(2:11)){datos[,i] <- cut(datos[,i],breaks = c(0, 4.99, 6.99, 8.99, 11.0),
labels = c("Suspenso", "Aprobado", "Notable", "Sobresaliente"))}
datos[,i]
```

Ilustración 5 cambio de variables cuantitativas a cualitativas

3. Obtenemos las tablas de probabilidad condicionada estimada con el clasificador de Naive Bayes.

```
b <- naiveBayes(NOTA.FINAL.JUNIO...0.10. ~ ., data = datos)
b
```

Ilustración 6 Naive Bayes

Las probabilidades a priori son:

Suspender	Aprobar	Notable	Sobresaliente
71,8%	24,2%	3,4%	0,7%

- Grupos

	A	B
Suspenso	54,2%	45,7%
Aprobado	36,1%	63,8%
Notable	40%	60%
sobresaliente	100%	0%

- Practica 1

	Suspens	aprobado	Notable	sobresaliente
Suspens	23,5%	15,2%	42,3%	18,8%
Aprobado	20%	14,2%	40%	25,7%
Notable	0%	20%	60%	20%
sobresaliente	0%	0%	0%	100%

- Practica 2

	Suspens	Aprobado	Notable	sobresaliente
Suspens	42,3%	20,3%	33,8%	3,3%
Aprobado	15,6%	25%	59,3%	0%
Notable	0%	0%	60%	40%
sobresaliente	0%	0%	0%	100%

- Practica 3

	Suspens	Aprobado	Notable	sobresaliente
Suspens	35,6%	17,8%	36,9%	9,5%
aprobado	14,2%	34,2%	42,8%	8,5%
Notable	0%	20%	60%	20%
sobresaliente	0%	0%	100%	0%

- Total practicas

	Suspens	Aprobado	Notable	sobresaliente
Suspens	55,3%	21,2%	21,2%	2,1%
aprobado	25%	33,3%	41,6%	0%
Notable	0%	40%	40%	20%
sobresaliente	0%	0%	100%	0%

- Cuestionario 1

	Suspens	Aprobado	Notable	sobresaliente
Suspens	53,6%	28,9%	17,3%	0%
aprobado	36,3%	42,4%	18,1%	3%
Notable	20%	20%	40%	20%
sobresaliente	0%	0%	100%	0%

- Cuestionario 2

	Suspens	Aprobado	Notable	sobresaliente
Suspens	59,1%	22,4%	16,3%	2%
aprobado	62%	27,5%	0%	10,3%
Notable	20%	20%	20%	40%
sobresaliente	0%	0%	100%	0%

- Cuestionario 3

	Suspens	Aprobado	Notable	sobresaliente
Suspens	66,7%	23,1%	10,2%	0%
aprobado	58,3%	12,5%	29,1%	0%
Notable	25%	25%	25%	25%
sobresaliente	0%	0%	100%	0%

- Total cuestionario

	Suspens	Aprobado	Notable	sobresaliente
Suspens	90%	7,5%	2,5%	0%
aprobado	69,4%	25%	5,5%	0%
Notable	40%	20%	20%	20%
sobresaliente	0%	0%	1005	0%

- Examen junio

	Suspens	Aprobado	Notable	sobresaliente
Suspens	95,6%	4,4%	0%	0%
aprobado	16,7%	75%	8,3%	0%
Notable	0%	20%	60%	20%
sobresaliente	0%	0%	0%	100%

Podemos ver que para los valores de Total Cuestionario, la predicción de suspens es totalmente condicionada a suspender los cuestionarios. Si sacas un sobresaliente tu nota se verá afectada positivamente.

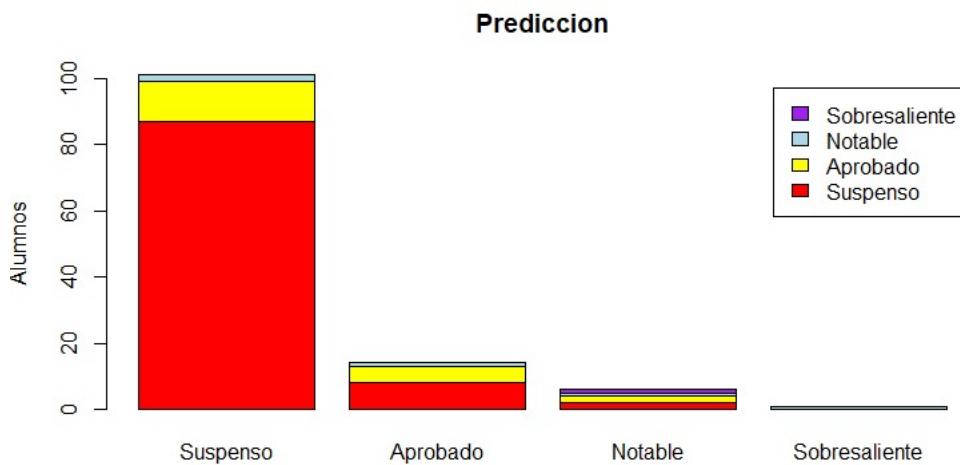


Ilustración 7 predicción total cuestionarios

Para la nota Total de las Prácticas, podemos ver que la predicción es muy difusa aun así, se puede ver que si se aprueban las prácticas es imposible suspender según la predicción.



4. Se analiza si es posible hacer una predicción mas ajustada al pasar de suspenso/aprobado a suspensos, aprobado, notable y sobresaliente.

```
tabla1 <- naiveBayes(datos$NOTA.FINAL.JUNIO...0.10. ~. , data = datos[,c(1:9,11)])
tabla1
table(predict(tabla1, datos[,c(1:9,11)]), datos$NOTA.FINAL.JUNIO...0.10.)
```

Ilustración 9 paso de suspenso/aprobado a suspensos, aprobado, notable y sobresaliente

Con todo estos datos obtenemos una predicción de que en junio habrá unas 107 suspensos y 42 aprobados, al igual que con la predicción de suspenso/aprobado, aunque con unos resultados desglosado diferentes.

La predicción cuenta con un margen de error, puesto que con naiveBayes obtenemos un resultado de:

- El número de alumnos que aprueben habiendo aprobado es de 21, una probabilidad de 14,09%
- El número de alumnos que aprueben habiendo suspendido es de 13, una probabilidad de 8,72%
- El número de alumnos que suspendan habiendo suspendido es de 94, una probabilidad de 63,09%
- El número de alumnos que suspendan habiendo aprobado es de 21, una probabilidad de 14,09%

Objetivo 3

- Al utilizar variables cuantitativas la predicción es peor que cuando utilizamos variables cuanlitativas, ya que no agrupa por intervalos para ver el numero de aprobados o suspensos.

0.1	0.13333	0.16667	0.23333	0.33333	0.334	0.38466	0.45	0.50183	0.52	0.5855	0.62333	0.65	0.69667
1	1	1	2	1	1	1	1	1	1	1	1	1	1
0.70333	0.718	0.76966	0.78566	0.80667	0.908	0.93	0.9325	0.93666	0.96667	0.98333	0.9845	0.99	1.00133
1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.05	1.11	1.12333	1.15133	1.16583	1.21	1.23667	1.27	1.282	1.3	1.35183	1.44	1.48	1.61
1	1	1	1	1	1	1	1	1	1	1	2	1	1
1.7	1.74	1.75	1.75916	1.79	1.832	1.8825	1.9	1.95	1.98333	2.11883	2.14333	2.1825	2.37333
1	1	2	1	1	1	1	1	1	1	1	1	1	1
2.46233	2.518	2.5325	2.64	2.719	2.7825	2.82667	2.89	2.95916	2.969	3.01333	3.01667	3.06	3.06966
1	1	1	1	1	1	1	1	1	1	1	1	1	1
3.08333	3.108	3.11	3.14	3.15	3.29333	3.32333	3.38333	3.46033	3.49	3.56666	3.59	3.68333	3.71667
1	1	1	1	1	1	1	1	1	1	1	1	1	1
3.7395	3.764	3.76667	3.85667	4.06333	4.21333	4.24916	4.40667	4.41333	4.49667	4.51	4.52466	4.53633	4.58066
1	1	1	1	1	1	1	1	1	1	1	1	1	1
4.60667	4.69333	4.71	4.72333	4.75	4.92333	5.01183	5.01233	5.05233	5.05616	5.14333	5.14366	5.16	5.17
1	1	1	1	1	1	1	1	1	1	1	1	1	1
5.179	5.23	5.27667	5.29366	5.31	5.3645	5.373	5.479	5.49	5.49366	5.51866	5.56	5.593	5.66733
1	1	1	1	1	1	1	1	1	1	1	1	1	1
5.87333	6.00066	6.02	6.12583	6.19667	6.2395	6.27616	6.28	6.35	6.40916	6.48533	6.72283	6.79333	6.89516
1	1	1	1	1	1	1	1	1	1	1	1	1	1
7.19516	7.479	7.48	7.85667	8.77916	9.38								
1	1	1	1	1	1								
>													
>													

Ilustración 10 tabla de predicciones

- Al pasar unas variables a cualitativas, observamos como la predicción es mayor.

Grupo

	A	B
Suspensos	54,2%	45,8%
Aprobados	38,1%	61,9%

Practica 1

	Suspensos	Aprobados
Suspensos	37,4%	21,6%
Aprobados	45,1%	18,4%

- Modificando el Excel anterior, introduciendo nuevas filas con los datos de los componentes del grupo, calculamos la probabilidad de aprobar de cada uno con Naive Bayes, obteniendo como resultado que solo 1 de 6 aprobará, por lo tanto la probabilidad de aprobar es de 16,67%

```

datos <- read.csv("Notas1.csv", sep=";", dec=",", header=T)

for(i in c(2:4)){(datos[,i])*3*5}
datos$Pr?.ctica.1..0.2.3. <- datos$Pr?.ctica.1..0.2.3.*(10/(2/3))
datos$Cuestionario.temas.1.y.2..0.10.<-datos$Cuestionario.temas.1.y.2..0.10.
datos$Cuestionario.tipo.test.temas.3.y.4...0.10.<-datos$Cuestionario.tipo.test.temas.3.y.4...0.10.

datos$Cuestionario.tipo.test.temas.5.y.6...0.10.<-NULL
datos$Pr?.ctica.2..0.2.3.<-NULL
datos$Pr?.ctica.3..0.2.3.<-NULL
datos$Total.cuestionarios..0.1.<-NULL
datos$EXAMEN.JUNIO..0.7.<-NULL
datos$NOTA.FINAL.SEPTIEMBRE..0.10. <- NULL
datos$EXAMEN.SEPTIEMBRE..0.7.<-NULL
datos

datos$NOTA.FINAL.JUNIO..0.10. <- cut(datos$NOTA.FINAL.JUNIO..0.10., breaks = c(0, 5, 10), labels = c("Suspensó", "Aprobado"))
datos$NOTA.FINAL.JUNIO..0.10.

notasFuturo <- naiveBayes(NOTA.FINAL.JUNIO..0.10.~ ., data = datos)
notasFuturo
predictions <- predict(notasFuturo, datos)
predictions
table(predictions, datos$NOTA.FINAL.JUNIO..0.10.)

```

Ilustración 11 proceso de los cálculos

```

predictions <- predict(notasFuturo, datos)
predictions
#> [1] Suspenso Suspenso Suspenso Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso
#> [16] Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Aprobado Aprobado Aprobado Aprobado Aprobado Suspenso Aprobado
#> [31] Suspenso Suspenso Aprobado Suspenso Aprobado
#> [46] Suspenso Suspenso Aprobado Aprobado Suspenso Aprobado Suspenso Suspenso Aprobado Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso
#> [61] Suspenso Suspenso Aprobado Suspenso Aprobado Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso
#> [76] Suspenso Aprobado Aprobado Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso
#> [91] Suspenso Aprobado Aprobado Suspenso Aprobado Suspenso Aprobado Aprobado Suspenso Aprobado Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso
#> [106] Suspenso Suspenso Suspenso Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Aprobado Suspenso Aprobado Suspenso Suspenso Aprobado Suspenso
#> [121] Suspenso Aprobado Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Aprobado Suspenso Suspenso Aprobado Suspenso Aprobado Suspenso
#> [136] Suspenso Suspenso Suspenso Aprobado Suspenso Suspenso Aprobado Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso Suspenso
#> [151] Suspenso Suspenso Suspenso Suspenso Suspenso
```

Ilustración 12 predicciones de los datos de los componentes

(los últimos seis datos son los de los componentes del grupo)