

Untitled

Iker Meneses Sales

2023-10-11

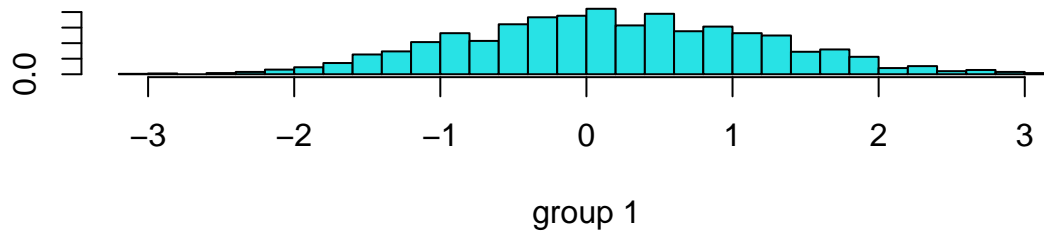
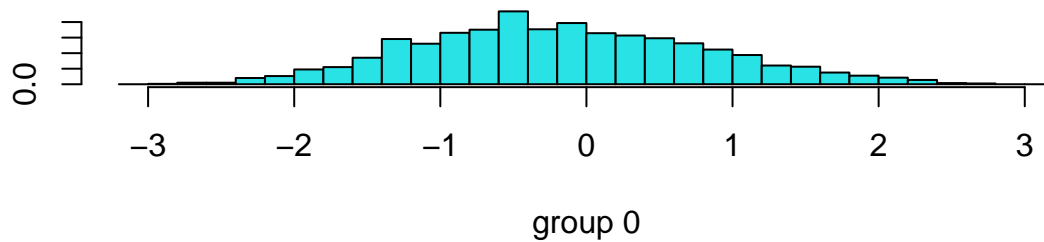
LDA (Linear Discriminant Analysis)

Para comenzar con los modelos discriminantes, se realizará en primer lugar un linear discriminant analysis (LDA) con el objetivo de intentar separar aquellos clientes que puedan tener dificultades de pago con aquellos solventes. Para ello, sin embargo, realizaremos un procedimiento ROSE (Random Over-Sampling Examples) con el objetivo de balancear nuestros datos evitando el over-fitting de nuestros datos y evitando una mala calibración de los algoritmos.

En primer lugar, estandarizamos los datos usando la función de R `scale()` para así hacer que todos las variables tengan el mismo peso. Una vez los datos han sido normalizados y tienen todas las variables el mismo peso en el modelo, aplicamos ROSE:

Seguidamente, se realiza una partición del dataset entre train y test con el objetivo de conseguir aproximaciones correctas para que puedan ser usadas, evitando el over-fitting. Para ello, se realizará una partición 80-20 de la base de datos:

Aplicando el discriminante lineal correspondiente, se obtienen los siguientes resultados en cuanto a la proyección de las observaciones se refiere:



A primera vista, se podría decir que el modelo separa de forma errónea, ya que los dos histogramas están sobrepuestos. Sin embargo, es necesario realizar la matriz de confusión:

Table 1: Matriz de confusión del conjunto test

	Realidad	
	No moroso	Potencial moroso
Predicción		
No moroso	318	202
Potencial moroso	209	271

Como ya se podía prever, los resultados son muy pobres: únicamente se ha obtenido una accuracy del 58.9%, segregado en una sensibilidad de un 60.3416% y una especificidad del 57.2939%.

Sin embargo, se sabe que el LDA puede presentar problemas en el momento en el que las variables no presentan normalidad o cuando las matrices de covarianzas son diferentes para cada grupo. Como ya se apreció en la descriptiva post-preprocessing, muchas de nuestras variables no presentaban normalidad, de forma que esto podría ser un problema de cara al uso del LDA. Es por eso por lo que se ha decidido realizar un QDA (Quadratic Discriminant Analysis) con el objetivo de corregir dichos problemas y mejorar la performance del LDA.

QDA

Así pues, repitiendo el procedimiento seguido anteriormente en el LDA, toca repetir los mismos pasos para este modelo. De esta forma, los resultados obtenidos son los siguientes:

Table 2: Matriz de confusión del conjunto test

	Realidad	
	No moroso	Potencial moroso
Predicción		
No moroso	313	201
Potencial moroso	214	272

Como se puede apreciar, los resultados obtenidos usando QDA son muy similares a los que se obtuvieron en el discriminante lineal. Si bien es cierto que la precisión del modelo es del 58.5%, algo peor que en LDA, parece que la especificidad (57.5053%) es superior en este modelo. Por contra, la sensibilidad es menor que en LDA (59.3928%)

En resumen, observando los resultados obtenidos, se puede afirmar que los dos modelos discriminantes presentan resultados muy pobres: es probable que el hecho de añadir posteriormente las variables categóricas acabe de hacer que se mejore de forma clara los resultados conseguidos hasta ahora.