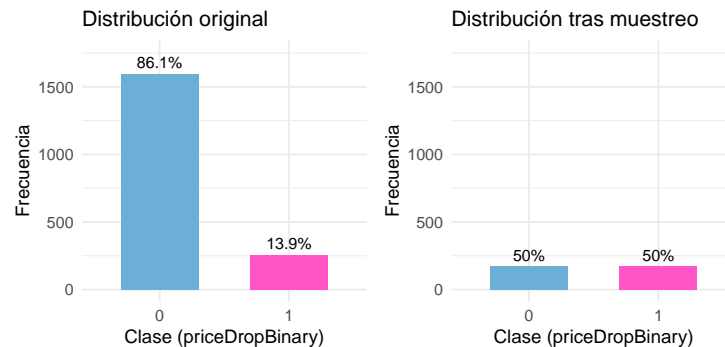


# PLS-DA RESUMIDO

Ana Valiente

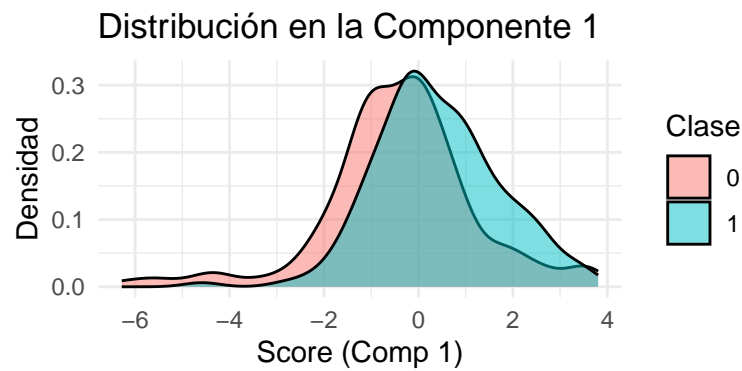
2025-06-03

En el análisis anterior, se ha probado que nuestra base de datos no cumple las condiciones de homocedasticidad ni normalidad. Por lo tanto, no es posible aplicar un Análisis Discriminante de Fisher para predecir qué inmuebles serán rebajados. En su lugar, emplearemos un PLS discriminante; en el que consideraremos como POSITIVOS las viviendas cuyo precio sea menor al precio original. Para ello, hemos convertido la variable primitiva priceAmountDrop en priceDropBinary (0 si el precio baja, 1 si se mantiene estable). En la primera fase del análisis, seleccionamos la mayoría de variables disponibles (incluyendo el cluster por servicios diseñado en análisis previos); descartando aquellas cuya información resultaba inservible o redundante. Para evitar el diseño de un modelo sesgado, escalamos las variables numéricas y dividimos los datos en los conjuntos de entrenamiento (70%) y test (30%). Además, realizamos un undersampling en la base de entrenamiento eliminando aleatoriamente parte de los pisos sin bajada de precio; puesto que había un fuerte desbalanceo que podía favorecer sistemáticamente las predicciones de la clase predominante.

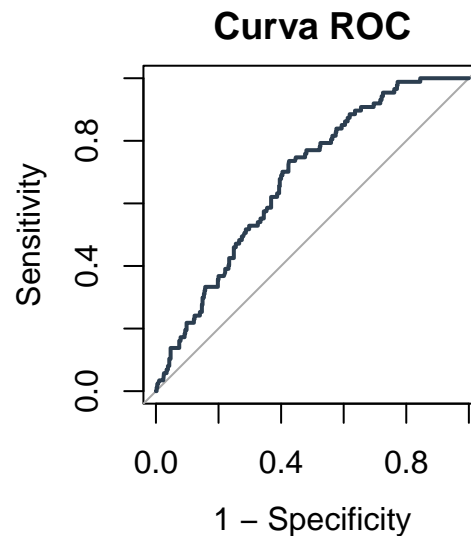


Seguidamente, propusimos el modelo inicial. Aplicamos la validación cruzada con 5 grupos o folds y 10 repeticiones hasta quedarnos con 1 única componente principal, minimizando así el error de predicción. Una vez aplicado el modelo al conjunto de test, maximizamos su exactitud balanceada de 0.59 a 0.632 variando el umbral óptimo. Seleccionamos esta métrica como medida de calidad porque tiene en cuenta los aciertos en ambas clases de pisos (0, 1), sin proporcionar resultados engañosos ni dejarse condicionar por el desbalanceo. Esta primera versión mostraba una capacidad predictiva aceptable, generando mejores resultados que los que se obtendrían por azar. Cabe recalcar que la bajada de precio de las viviendas está muy condicionada por el factor humano y las condiciones individuales de cada propietario. Por lo tanto, podríamos considerarla una variable de naturaleza impredecible o, al menos, con un comportamiento difícil de modelar. A continuación, nos propusimos incrementar la eficacia mediante un test de independencia. El objetivo era proponer un nuevo modelo formado por las variables que presentaran diferencias significativas entre clases. Así, descartamos la inclusión de variables sin capacidad predictiva, que únicamente generan ruido y disminuyen la fiabilidad de las predicciones. A las variables numéricas les aplicamos el test t y a las categóricas el test chi-cuadrada o, en caso de muestras pequeñas, la prueba de Fisher. Todo el proceso descrito hasta este punto está detallado en el anexo del PLS-DA. Una vez realizada la selección, volvimos a dividir, escalar y reducir la nueva base de datos y generamos el segundo modelo; compuesto nuevamente por una sola componente principal.

## Componentes óptimos: 1



Aunque la distribución obtenida era relativamente uniforme, maximizamos una vez más la exactitud balanceada a través de la curva ROC:



## Umbral óptimo para máxima Balanced Accuracy: 0.474

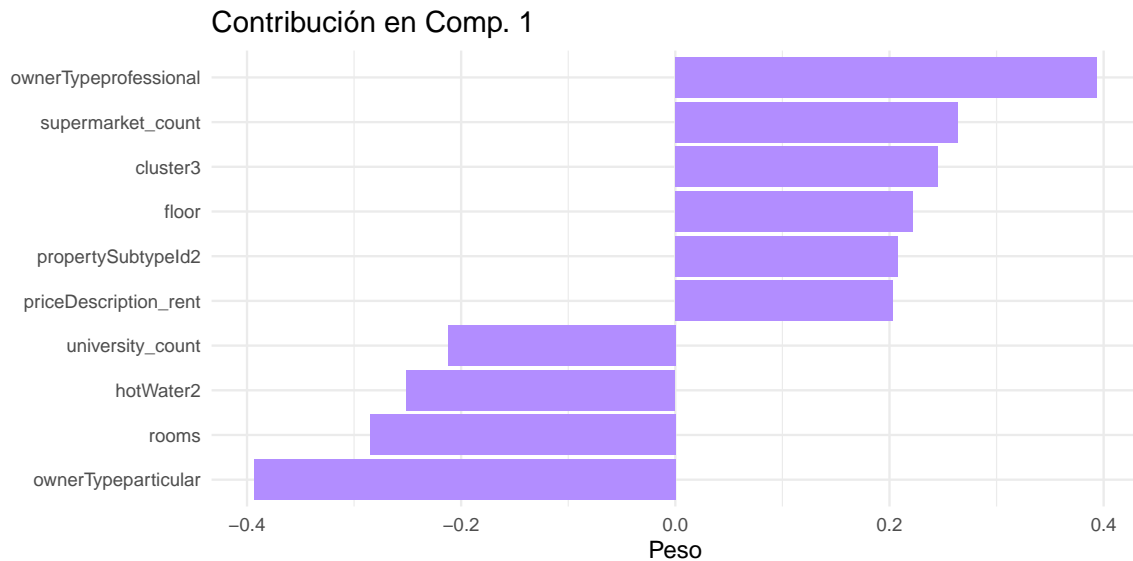
```
##      Real
## Predicho  0  1
##      0 270 23
##      1 200 64
```

## Accuracy: 0.6

## Balanced Accuracy: 0.655

Finalmente, conseguimos un modelo que predice las bajadas de precio de los inmuebles con un 65% de exactitud balanceada. Esto es, cada 100 predicciones se consigue una media razonable de 65 recalls. Tal como hemos indicado en el inicio del análisis, esta variable está fuertemente condicionada por factores

humanos: decisiones personales de propietarios, estrategias de agencias inmobiliarias, urgencia de venta o negociación, entre otros elementos difíciles de cuantificar. Por tanto, ningún modelo podrá predecirla con exactitud absoluta, y siempre quedarán fuera variables imposibles de medir que condicionarán los resultados. Aun así, cualquier herramienta que logre predecir las fluctuaciones en los precios, aunque sea parcialmente, es altamente valiosa; ya que permite orientar decisiones y ofrecer una ventaja estratégica en el análisis de mercado inmobiliario. Una vez elaborado el modelo definitivo, estudiaremos qué variables tienen más peso en la componente principal para entender qué características hacen a un piso más propenso a ser rebajado.



Resalta claramente que los propietarios profesionales (inmobiliarias, plataformas online, entidades bancarias...) son más propensos a reducir el precio que los propietarios particulares. Por otra parte, si nos centramos en la ubicación, en las zonas de Ruzafa, Benimaclet y Jesús (cluster3) las viviendas son más propensas a las rebajas. Asimismo, cuanto mayor sea el precio medio del alquiler en esta ubicación, más probable será que estos precios bajen. La cantidad de supermercados próximos también contribuye positivamente a dicha probabilidad. Por el contrario, precios de pisos en zonas universitarias son más estables. Respecto a las características del piso per se, las plantas altas tienen más probabilidad de ser rebajadas. En cambio, si tiene muchas habitaciones y/o calentador eléctrico (hotWater2) tiende a experimentar menos variaciones; especialmente en el caso de los apartamentos (propertySubtype2).