Altman Score

Alejandro García de León

2025-12-27

Modelo matemático

Una transformación lineal que es idempotente, es decir $T \circ T = T$ y normal es decir $T \circ T^* = T^* \circ T$ (conmuta con su operador adjunto) es una proyección ortogonal y el rango es el complemento ortogonal del dominio, es decir el conjunto de todos aquellos elementos cuyo producto interior con los elementos del dominio es 0 o bien el Kernel del producto interior en una coordenada. Ahora bien $Id - T$ es la otra proyección y sumadas dan la identidad. Entonces el vector proyección es el mas cercano al proyectado en la distancia inducida por el producto interior. Cuando estamos en espacios de Lebesgue $L^2(P)$ esta norma tiene un producto interior y su proyección viene dada por los coeficientes por mínimos cuadrados. Ahora bien puede aplicarse esa proyección a una n-eada y eso generaría un modelo de regresión lineal. Esa proyección es $\hat{y} = \beta\vec{x}$; si quisieramos explicar un vector y en terminos de otros $\vec{x} = (x_1, \ldots, x_m)$ cada uno conn coordenadas se calcula $N = \int(y - \hat{y})^2$ y se escoje la proyección β como la que minimiza N. Este es el razonamiento dentro del análisis Z-Altman.

Base utilizada

```r
library(readr)
clientes_aseguradora_sintetico <- read_csv("C:/Users/Alejandro/Downloads/WPy64-38123/notebooks/clientes_
```

```
## Rows: 200000 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): nombre
## dbl (8): cliente_id, edad, ocupacion_seguridad, rango_salarial, calificacion...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
head(clientes_aseguradora_sintetico,7)
```

```
## # A tibble: 7 x 9
##   cliente_id nombre            edad ocupacion_seguridad rango_salarial
##        <dbl> <chr>            <dbl>               <dbl>          <dbl>
## 1          1 Ana Rodríguez       56                   2         20591.
## 2          2 Lucía Rodríguez     69                   2         61539.
## 3          3 Laura Rodríguez     46                   1         43641.
## 4          4 Juan Hernández      32                   2         20385.
## 5          5 Luis Pérez          60                   4         19751.
```

```
## 6          6 José Hernández      25                    4         50762.
## 7          7 Carlos Rodríguez    38                    3         31858.
## # i 4 more variables: calificacion_crediticia <dbl>, sexo <dbl>,
## #   endeudamiento_ingreso <dbl>, tier_cumplimiento <dbl>
```

La base fue realizada con datos simulados para ejemplificar las técnicas de análisis multivariado aquí abordadas.

Modelo matemático

A continuación se usa la notación $<;>$ para denotar un producto interior.

Dentro de la regresión lineal tenemos la matriz sombrerro H, esta es tal que $HY = \beta \cdot X = \hat{Y}$.

Esto se obtiene de calcular la derivada parcial del error cuadrático medio

$$\frac{\partial ECM}{\partial \beta} = \frac{\partial}{\partial \beta}(Y - X\beta)^t(Y - X\beta) = -2X^t(Y - X\beta)$$

Igualando a 0 y depejando llegamos a que $X^tX\beta = X^tY$ es decir $\beta = (X^tX)^{-1}X^tY$; ya que $HY = \hat{Y} = X\beta$ entonces $H = X(X^tX)^{-1}X^t$.

Por tanto sabemos que para resolver mínimos cuadrados hay que resolver $Cov(X)\vec{\beta} = X^tX\vec{\beta} = X^tY$ (ecuaciones Yule-Walker), donde el lado derecho es el vector formado por la covarianza de Y y el último elemento del vector X en primer lugar, hasta llegar al primero de X. $\beta_0 = (1 - \sum \beta_i)\mu$.

H es la proyección ortogonal sobre el espacio generado por las columnas de la matriz X. Entonces $HY \in Span(X)^{\perp}$ y por ende $< Y - HY; X >= 0$ de tal suerte que $< Y; X > - < HY; X >= 0$ es decir $X^tY =< Y; X >=< HY; X >= X^tHY$ para cada $Y \in Span(X)$. Por ende nuevamente para que esto sea verdad es necesario que $H = X(X^tX)^{-1}X^t$.

Ahora tomando las covarianzas $T = X^tHX = X^tX$, estas pueden descomponerse de la forma $T = W + B$, donde W es una matriz que representa las covarianzas de elementos de un mismo grupo mientras que B las de grupos distintos. Si se pide la máxima homogeneidad posible es decir $a^t\ W\ a = 1$ donde a es la proyección buscada, entonces se llega a que $a^t\ T\ a = 1 + a^t\ B\ a$.

Se puede definir el lagrangiano:

$$L(a, \lambda) = a^t\ B\ a - \lambda(a^t\ W\ a - 1)$$

Derivando queda $2Ba - \lambda(2Wa) = 0$ implicando que $Ba = \lambda Wa$; de aquí $Ba = \lambda(Wa) \Leftrightarrow W^{-1}Ba = \lambda a$ esto implica que a es un eigenvector de $W^{-1}B$ que optimiza el cociente

$$\frac{a^tBa}{a^tWa}$$

y que satisface la restricción.

Aplicación del análisis anterior a un caso real de selección de cartera de una aseguradora

En matemáticas dado un vector de observaciones Y que en este caso se interpreta como un vector multidimensional que contiene las siguientes variables de un cliente escogido al azar de una aseguradora llamada **Aseguradora A** se interpreta como una selección aleatoria puede aplicarse una función

$$f : R^n \longrightarrow R$$

a cada cliente para medir cierto grado de alguna cualidad tomando en cuenta:

```r
colnames(clientes_aseguradora_sintetico )
```

```
## [1] "cliente_id"              "nombre"
## [3] "edad"                    "ocupacion_seguridad"
## [5] "rango_salarial"          "calificacion_crediticia"
## [7] "sexo"                    "endeudamiento_ingreso"
## [9] "tier_cumplimiento"
```

En nuestro caso quisieramos medir que tan bueno es un cliente de nuestra cartera para ubicarlo con mayor o menor probabilidad de incumplimiento.

La función distancia a utilizar será el error cuadrático medio.

Como medida para un cliente bueno, se tomará la distancia a elementos mas parecidos con las siguientes escalas:

Se dice que un cliente es

```r
#4 grupos
#solo 6 variables
#edad
#ocupacion_seguridad
#rango_salarial
#(Inversamente proporcional)
#calificacion_crediticia
#endeudamiento_ingreso
#tier_cumplimiento
X = clientes_aseguradora_sintetico

X = X[,c(3:6,8:9)]
X = as.matrix(X)
#bueno
c_1 = c(35, 5,80000,1,0,0)
#segundo mejor
c_2 = c(40, 4,30000,2,.3,1)
#tercer mejor
c_3 = c(50, 3,10000,3,.7,2)
#malo
c_4 = c(60, 2,8000,4,.8,3)
#Calcular la media de cada variable
s = c(1:ncol(X))
for(i in 1:ncol(X)){
s[i] = mean(X[,i])
}
#edad
#ocupacion_seguridad
#rango_salarial
#(Inversamente proporcional)
#calificacion_crediticia
#endeudamiento_ingreso
#tier_cumplimiento
```

Se construye el error cuadrático medio respecto de estos cuatro grupos

```r
n = nrow(X)
r1 = t(t(c_1-s))%*%t(c_1-s)
r1 = n*r1
r2 = t(t(c_2-s))%*%t(c_2-s)
r2 = n*r2
r3 = t(t(c_3-s))%*%t(c_3-s)
r3 = n*r3
r4 = t(t(c_4-s))%*%t(c_4-s)
r4 = n*r4
B = r1 + r2 + r3 + r4
B
```

```
##               [,1]          [,2]          [,3]          [,4]          [,5]
## [1,]      73796720    -8597099.0 -1.862335e+11  8.596718e+06  2.318140e+06
## [2,]      -8597099     1201804.0  1.992860e+10 -1.201013e+06 -2.137834e+05
## [3,] -186233486824 19928597026.2  7.403935e+14 -1.994300e+10 -8.244673e+09
## [4,]       8596718    -1201012.5 -1.994300e+10  1.200224e+06  2.140431e+05
## [5,]       2318140     -213783.4 -8.244673e+09  2.140431e+05  1.037272e+05
## [6,]       8789284    -1601229.4 -1.266189e+10  1.598871e+06  8.272265e+04
##               [,6]
## [1,]  8.789284e+06
## [2,] -1.601229e+06
## [3,] -1.266189e+10
## [4,]  1.598871e+06
## [5,]  8.272265e+04
## [6,]  2.791227e+06
```

Estas son las sumas de cuadrados de grupos exteriores.

Las de grupos interiores serán la covarianza entre columnas

```r
W = cov(X)
```

La idea del análisis discriminante de Fisher es la siguiente: Resolver el problema lineal

$$max(a^t \ B \ a)$$

s.a a la restricción

$$a^t \ W \ a = 1$$

La solución al mismo esta dada por a igual a eigenvector de $W^{-1}B$ que correspomde al eigenvalor más grande de esta matriz.

Calculamos los eigenvectores

```r
x = solve(W)
x = x%*%B
valores = eigen(x)[[1]]
valores
```

```
## [1] 6.279246e+06+0.000000e+00i 1.330801e+06+0.000000e+00i
## [3] 1.331936e+05+0.000000e+00i 6.834490e+03+0.000000e+00i
## [5] 4.867999e-11+1.265122e-10i 4.867999e-11-1.265122e-10i
```

Como son reales estos eigenvalores nos quedamos solo con la parte real:

```
valores = Re(valores)
valores
```

```
## [1] 6.279246e+06 1.330801e+06 1.331936e+05 6.834490e+03 4.867999e-11
## [6] 4.867999e-11
```

el más grande es

```
max(valores)
```

```
## [1] 6279246
```

Cuyo eigenvector correspondiente es

```
eigenvectores = eigen(x)[[2]]
eigenvectores = eigenvectores[,1]
eigenvectores  =Re(eigenvectores)
eigenvectores
```

```
## [1]  2.243585e-03 -3.337823e-02 -4.454371e-06  3.340687e-02  9.988156e-01
## [6]  1.150142e-02
```

Esta es la proyección buscada. Para aplicar esta regla se resta al cliente que se desee clasificar el valor representativo de cada grupo y se le aplica la función. El que sea más pequeño será el grupo que pertenecerá este cliente.

Ejemplificación de lo anterior

Suponiendo llega un cliente con las siguientes características: 70 años, buena ocupación de grado 5, un salario de $50,000, endeudado al 50% y que nunca ha incumplido sus créditos previos (tier 0).

```
h = c(70, 5,50000,1,.5,0)
abs(sum(eigenvectores*(c_1-h)))
```

```
## [1] 0.7115644
```

```
abs(sum(eigenvectores*(c_2-h)))
```

```
## [1] 0.09969673
```

```
abs(sum(eigenvectores*(c_3-h)))
```

```
## [1] 0.4896393
```

```r
abs(sum(eigenvectores*(c_4-h)))
```

```
## [1] 0.699152
```

Entonces el segundo grupo es aquel al que pertenece este cliente, es decir es un cliente de segundo nivel.