

Uso de LDA en MATLAB

Ulises Jiménez Guerrero

7 de abril de 2025

Como continuación de la actividad realizada anteriormente, sobre el uso de PCA en Matlab, se desarrolló esta actividad ejemplificando el uso del análisis discriminante lineal (LDA) en este mismo entorno de programación. Nuevamente se utilizó la base de datos *wine_quality*, que mide 11 variables en relación con la calidad de un vino rojo [Cortez et al., 2009].

Para el uso de esta técnica se recurrió a la función *LDA* [Heris, 2015], que no es nativa a Matlab. Esta función toma como entrada la base de datos y las clases correspondientes a cada observación. Devuelve tres elementos: los datos transformados de acuerdo a LDA, la matriz de eigenvectores y los eigenvalores correspondientes. Dada la forma en que está diseñado el algoritmo, este intentará transformar los datos de manera que cada clase se encuentre distribuida en clústers bien diferenciados.

A continuación se muestra el script desarrollado para el análisis en Matlab:

```
% Lectura de los datos
wine_quality = readtable("winequality/winequality-red.csv");
quality = categorical(wine_quality.quality);
data = table2array(wine_quality);
data = data(:, 1:end-1);

% Determinacion del numero de clases presente
classes = unique(quality);
num_classes = length(classes);

% Aplicacion de LDA
[Y, W, lambda] = LDA(data, quality);

% Graficacion de los datos
graph_data = Y(:, 1:3);
for i=1:num_classes
    current_data = graph_data(quality==classes(i), :);
    scatter3(current_data(:,1), current_data(:,2), current_data(:,3), ...
            10, 'filled')
    hold on
end
title('LDA, 3 dimensiones')
legend(classes)
hold off
```

Se obtuvo la siguiente matriz de vectores, junto a los eigenvalores correspondientes:

$$\begin{bmatrix} -0.001 & -0.001 & -0.000 & 0.002 & -0.000 & -0.001 & 0.002 & -0.001 & 0.001 & 0.001 \\ -0.001 & & & & & & & & & \\ 0.016 & -0.010 & -0.006 & -0.005 & 0.006 & 0.001 & 0.004 & -0.002 & 0.001 & 0.002 \\ -0.001 & & & & & & & & & \\ 0.002 & -0.004 & -0.008 & -0.002 & 0.006 & 0.022 & -0.007 & -0.004 & 0.001 & -0.011 \\ 0.003 & & & & & & & & & \\ -0.001 & -0.001 & -0.001 & 0.001 & -0.001 & 0.001 & 0.001 & -0.000 & -0.000 & 0.000 \\ -0.000 & & & & & & & & & \\ 0.034 & -0.007 & 0.011 & 0.015 & 0.017 & -0.050 & 0.015 & 0.012 & -0.008 & -0.060 \\ 0.006 & & & & & & & & & \\ -0.000 & 0.000 & 0.000 & -0.000 & 0.000 & 0.000 & 0.000 & -0.000 & 0.000 & -0.000 \\ -0.000 & & & & & & & & & \\ 0.000 & 0.000 & -0.000 & 0.000 & -0.000 & -0.000 & 0.000 & 0.000 & -0.000 & -0.000 \\ 0.000 & & & & & & & & & \\ 1.000 & 1.000 & 1.000 & -1.000 & 1.000 & -1.000 & -1.000 & 1.000 & -1.000 & 1.000 \\ 1.000 & & & & & & & & & \\ 0.002 & -0.010 & 0.002 & 0.015 & -0.006 & -0.031 & 0.021 & -0.000 & 0.004 & -0.022 \\ 0.002 & & & & & & & & & \\ -0.019 & -0.002 & -0.001 & -0.000 & -0.002 & -0.012 & 0.007 & -0.001 & -0.004 & 0.000 \\ -0.001 & & & & & & & & & \\ -0.005 & 0.001 & 0.000 & -0.001 & 0.001 & -0.004 & 0.001 & 0.001 & -0.001 & 0.000 \\ 0.001 & & & & & & & & & \end{bmatrix}$$

$$\begin{bmatrix} 0.63189 \\ 0.07644 \\ 0.02479 \\ 0.00642 \\ 0.00417 \\ 4.58928e-16 \\ 9.85301e-17 \\ -1.0592e-15 \\ -1.8812e-15 \\ -6.1340e-15 \\ -1.0936e-14 \end{bmatrix}$$

El algoritmo debería descartar de manera automática las dimensiones que no aporten datos sobre la clasificación de las variables. Dados los resultados, todas las variables aportan información, aunque los últimos 5 componentes son tan pequeños que se pueden despreciar. Sin embargo, resulta imposible visualizar 5 dimensiones. Por lo tanto, se realiza una visualización tomando los primeros 3 componentes, y se muestra junto a una visualización del PCA con el mismo número de dimensiones.

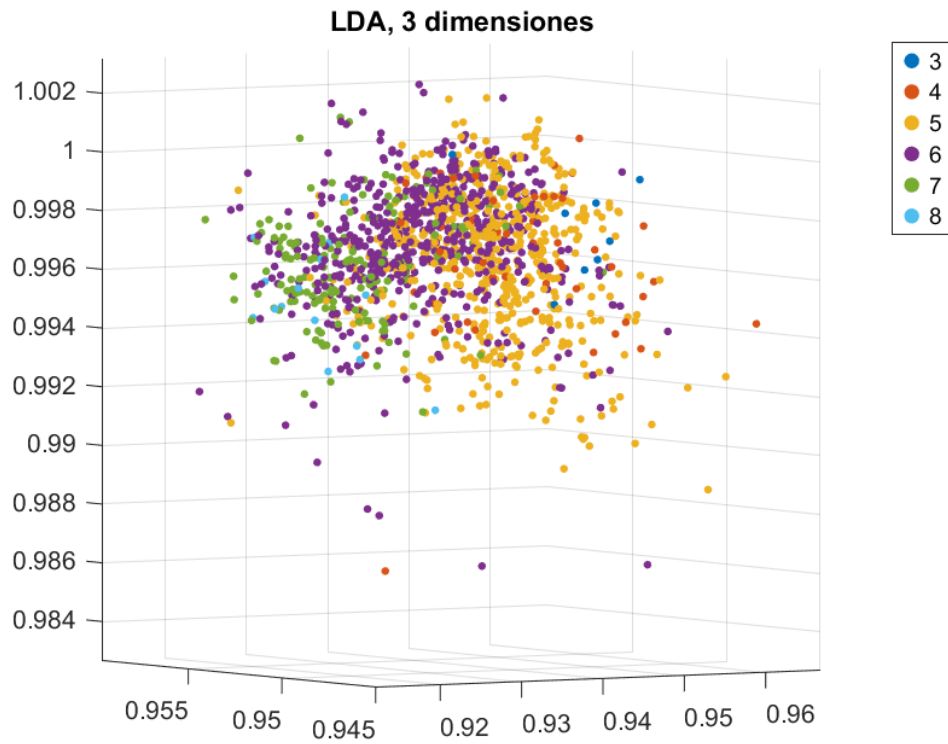


Figura 1: Primeras 3 dimensiones del análisis LDA

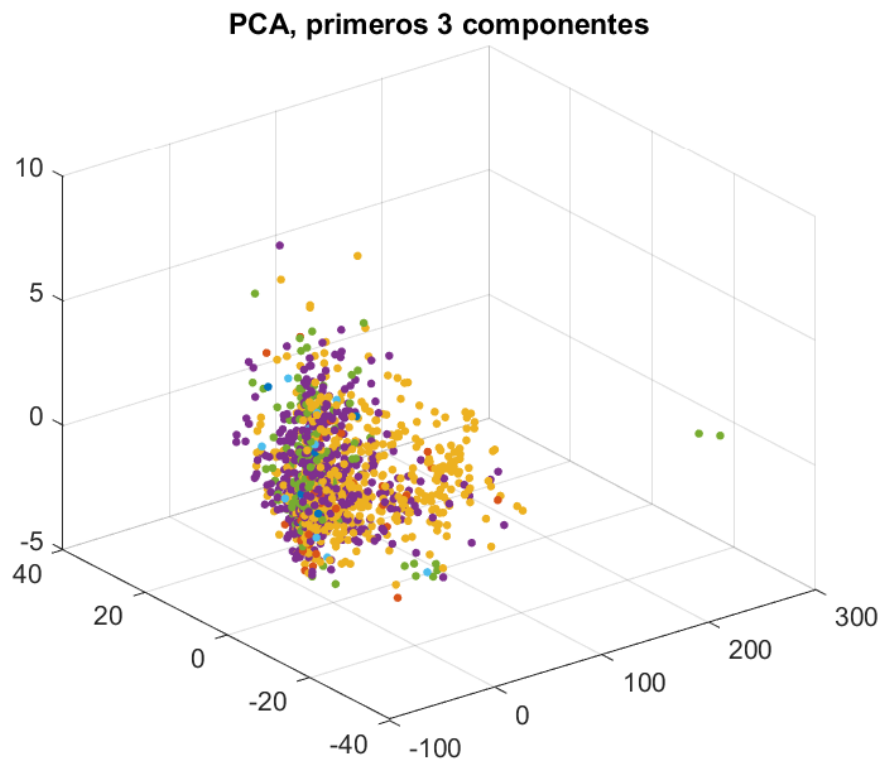


Figura 2: Primeras 3 componentes de PCA

No se logra una distinción clara de los datos en base a la calidad del vino. Esto se debe a diversos factores, como el que solo se utilizan 3 dimensiones para la representación, en vez de las 5 recomendadas por el algoritmo. Otra razón es que los datos pueden no encontrarse separados linealmente, lo que ocasiona problemas para este método. Sin embargo, al comparar con PCA, se observa una mejoría a la hora de representar las clases. En LDA se observan clústers superpuestos, mientras que no se observa ningún clúster para PCA.

Referencias

- [Cortez et al., 2009] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Wine Quality. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C56S3T>.
- [Heris, 2015] Heris, M. K. (2015). Linear discriminant analysis (lda) in matlab. <https://yarpiz.com/430/ypml114-linear-discriminant-analysis>.