

# Práctica 1:

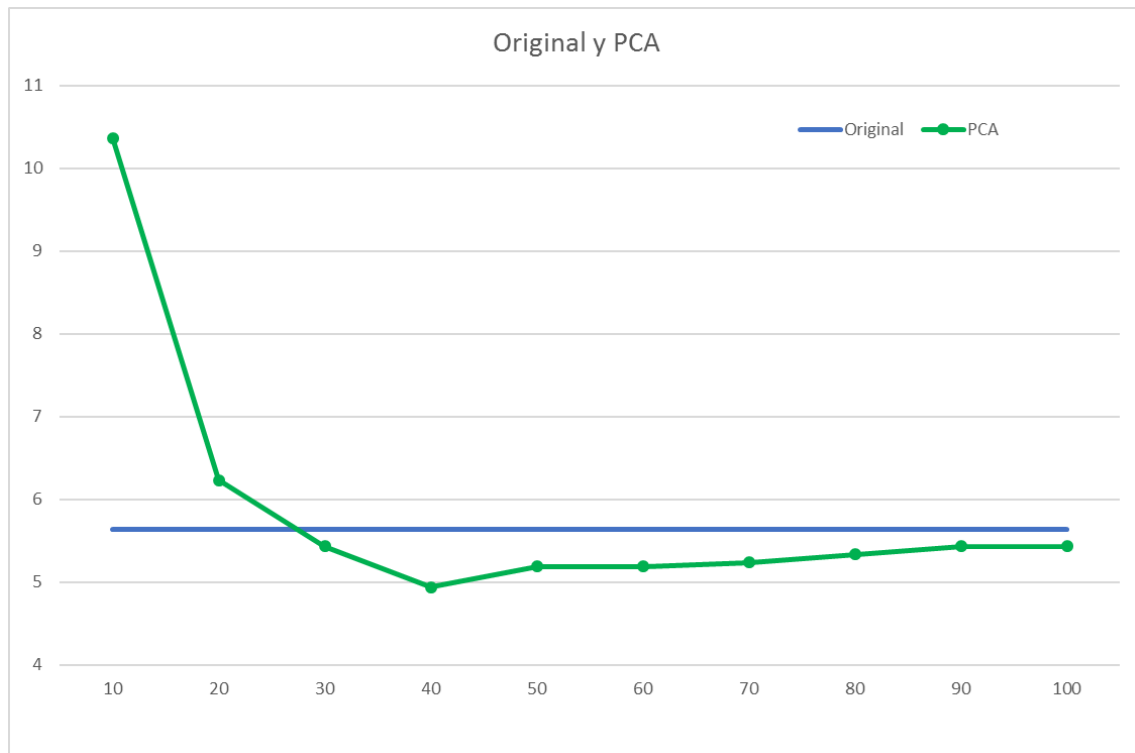
## Aplicación de PCA y LDA a OCR

Ramon Ruiz Dolz

Salvador Marti Roman

3CO21

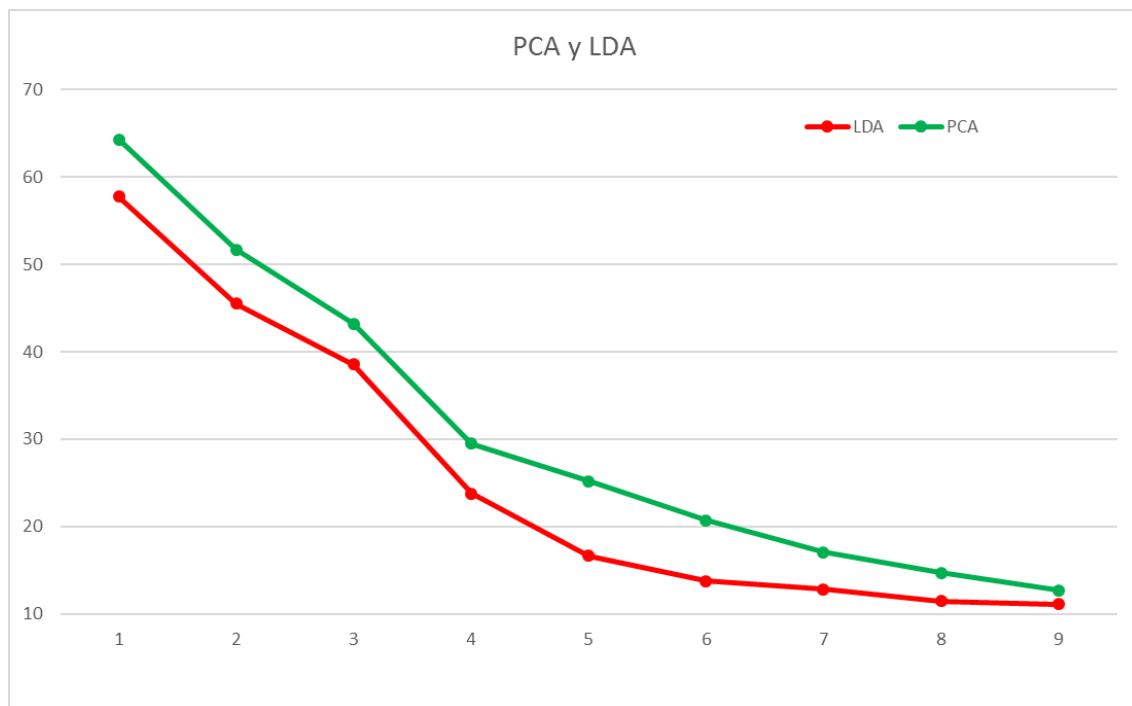
## *Espacio original y PCA*



Esta primera grafica representa la diferencia del % de error (eje vertical) en función de la dimensionalidad (eje horizontal) entre el conjunto de datos original y el conjunto de datos aplicando PCA. Como podemos observar, existe un punto a partir del cual el error empieza a aumentar, este es el 40 por lo tanto, sabemos que al reducir a más de 40 la dimensionalidad comenzaremos a tener problemas para separar y clasificar los datos de manera precisa. Por lo tanto, podemos concluir que con PCA no es más conveniente reducir lo máximo posible la dimensionalidad sin tener nada más en cuenta, si no decrementar al máximo esta siempre que no perdamos separabilidad. En este caso, reducir la dimensionalidad a 40 nos proporcionaría el menor error posible entre PCA y la original, como podemos observar. El % de error para cada dimensión en PCA es el siguiente: (En el espacio original es 5.63%)

Dim	Error
10	10.368 %
20	6.231 %
30	5.433 %
40	4.935 %
50	5.184 %
60	5.184 %
70	5.234 %
80	5.333 %
90	5.433 %
100	5.433 %

## PCA y LDA



Finalmente, en esta segunda gráfica vamos a comparar el % de error sobre un conjunto de datos aplicando PCA y LDA en el intervalo  $[1:C-1]$  siendo  $C = 10$ . Como en la anterior gráfica, el eje vertical es el % de error y el horizontal la dimensionalidad. Con esta gráfica, podemos observar que, para este intervalo de valores, es decir, el intervalo de valores en el que es aplicable LDA siempre será mejor LDA que PCA, ya que PCA se suele proyectar a mayor dimensionalidad y posteriormente LDA a una de estas. Mientras que LDA utiliza muestras etiquetadas, PCA no, y al reducir tanto la dimensionalidad de la matriz, se hace prácticamente imposible separar estas, obteniendo un alto % de error. Los valores del % de error para cada dimensión mediante el uso de PCA o LDA quedan reflejados en la tabla inferior.

<u>Dim</u>	<u>LDA(%)</u>	<u>PCA(%)</u>
1	57.776	64.257
2	45.663	51.694
3	38.285	43.220
4	23.180	29.461
5	17.597	25.174
6	14.007	20.737
7	12.063	17.048
8	11.216	14.705
9	10.667	12.711

Como podemos observar siempre es ligeramente superior en PCA, por eso a estas dimensiones siempre se recomienda aplicar LDA, a poder ser tras haber realizado una reducción con PCA a una dimensión mayor que  $C-1$ .