**Clusters Óptimos Jerárquicos:**

Se calcularon los índices cofeneticos para las combinaciones….

Si bien el mejor índice cofenético lo obtuvo la combinación de método y distancia: average y euclidean, cuando ejecutamos el algoritmo y probamos distintos puntos de corte, el 80 % de todos los géneros se clasificaban en un mismo cluster, quedando el resto de los clústers con muy pocos casos. Esto mismo nos ocurrió con muchas de las combinaciones medidas con el coeficiente cofenético. Decidimos guiarnos no sólo por el índice, sino también por los dendrogramas generados y las matrices de confusion.

Así fue que la mejor opción en casi todos los subconjuntos, fue el método “ward”, con distancia euclidea. Salvo por dos casos en que aplicamos el método “complete”.

Para elegir la cantidad de clústers óptimos, calculamos el índice de rand para rangos de maxclust desde 2 a 15 y juntamente con la visualización de los dendogramas, elegimos el número cluster óptimo para cada subconjunto, teniendo en cuenta también el sentido común.

**Punto 6 3 mejores géneros y 3 peores géneros**

Para poder determinar los mejores y peores géneros clasificados a lo largo de todas las experiencias, utilizamos la medida de pureza.

Por cada experiencia, calculamos la matriz de confusión comparando los géneros originales y los clústers óptimos recomendados según cada algoritmo. Luego, por cada género original, calculamos el índice de pureza,

Esta información la obtuvimos para cada subconjunto de datos y para cada algoritmo, contando 45 observaciones para los géneros mejor clasificados y 45 para los peores clasificados. De ahí obtuvimos que los mejores clasificados fueron:

mal:

jazz

singer-songwriter

drum-and-bass

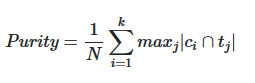
bien:

death-metal

opera

classical

Donde *N*= cantidad total de elements, *k* = número de clusters, *ci* es un cluster en *C*, y *tj* es la clasificación que tiene la máxima cantidad para el cluster *ci*



bibliografía :

[Christopher D. Manning](http://nlp.stanford.edu/%7Emanning/), [Prabhakar Raghavan](http://theory.stanford.edu/~pragh/) and [Hinrich Schütze](http://www.cis.uni-muenchen.de/personen/professoren/schuetze/), *Introduction to Information Retrieval*, Cambridge University Press. 2008.