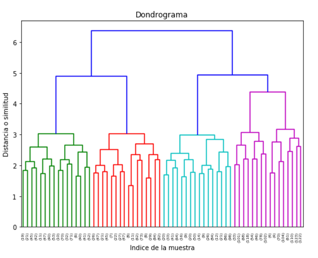


Nombre: **Israel Castillo Herrera. If708348**

Lea detenidamente los reactivos y responda con claridad. Si se requiere hacer uso de más hojas para la realización de cálculos, es necesario que se adjunten a este cuando se haga entrega del examen.

1. (3 puntos) En un experimento se lograron identificar 6 variables importantes que eran las que determinaban el comportamiento de este. Después de hacer muchos experimentos se lograron recoger las muestras de diferentes condiciones de trabajo y se encuentran en el archivo *dataset\_1\_b.csv*.  Determine cuantos grupos o patrones se encuentran en los datos recopilados y justifique su respuesta con código, figuras o mediciones.



**Se aplicó un clustering jerárquico primero y se visualizó el dendograma para determinar posibles separaciones de los grupos. En el dendograma, nos aparece que son posibles 5 grupos mediante este método.**

Z= hierarchy.linkage(data,metric= 'euclidean', method= 'complete')

plt.figure(figsize=(8,6))

dn2 = hierarchy.dendrogram(Z,

                         truncate\_mode='level',

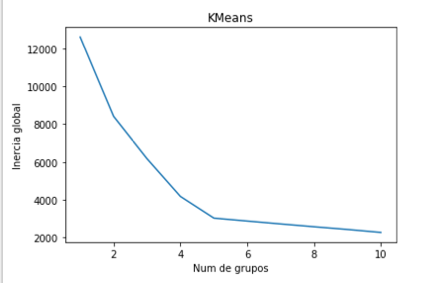
                         p=5)

plt.title('Dondrograma')

plt.xlabel('Indice de la muestra')

plt.ylabel('Distancia o similitud')

plt.show()



**En este otro algoritmo, generamos con el método de K-Means la separación de grupos que nos dio la gráfica para observar que el grupo óptimo son de 2 a primera vista. Pero para una inercia menor, el segundo codo se forma en 5 grupos, y siendo secundado por el dendograma.**

inercias = np.zeros(10)

for k in np.arange(10)+1:

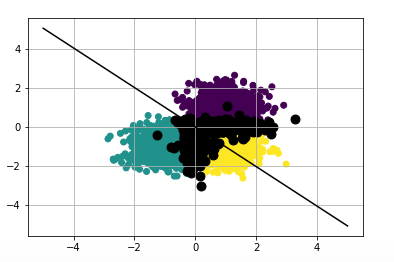
   model = KMeans(n\_clusters=k,init = 'random')

   model = model.fit(data)

   inercias[k-1] = model.inertia\_

1. (3 puntos) En una fábrica de piezas automotrices detectaron que la calidad de las piezas que se producían podía evaluarse realizando las medidas de dos equipos dentro de la fábrica. Las categorías de calidad definidas son “Bueno”, “Regular”, “Malo”. Después de realizar varias muestran lograron crear una base de datos con algunas muestras ya clasificadas (dataset\_2\_b.csv).  Realicé el diseño de un clasificador que logré identificar las categorías especificadas en la base de datos. Entregue el código que se utilizó para el diseño del clasificador, justifique y valide su respuesta por medio de medidas y gráficas.

**Con la estimación de nuestro modelo, nos quedan centrados los puntos negros, que serían nuestro vector soporte. Para respaldar lo que hicimos fue sacar el accuracy, precision\_score y nuestro recall\_score, con un valor por encima de un 90%  por lo que podemos decir que es seguro.**

https://lh5.googleusercontent.com/6-OiZwzrZWplQ6ey1mPPbkJmgybmamaG-dxQ0qh7rdpYbpeSedkHBg4xEM_V97HI-uah7TqbQpIqWqErHDq8CNlzKGK2SVJB9_BCG3BL1M7lwmFcP3qjWdKLDFRmhLmqxEISKvCy

#EJERCICIO 2

#%%

data = pd.read\_csv('../Data/dataset\_2\_b.csv')

#%%

dqr\_data = mylib.dqr(data)

def replace\_text(x,to\_replace,replacement):

   try:

       x = x.replace(to\_replace,replacement)

   except:

       pass

   return x

#%%

modelo = svm.SVC(kernel = 'linear')

y = data.Class

y = replace\_text(y,'Malo',0)

y = replace\_text(y,'Regular',1)

y = replace\_text(y,'Bueno',2)

x = pd.DataFrame(data.V1).join(data.V2)

modelo.fit(x,y)

yhat = modelo.predict(x)

yhat = pd.DataFrame(yhat)

w = modelo.coef\_[0]

m= -w[0]/w[1]

b = -modelo.intercept\_[0]/w[1]

xx = np.linspace(-5,5)

yy = m\*xx+b

vs = modelo.support\_vectors\_

Xs = modelo.support\_vectors\_[0]

values\_y1 = m\*xx+(Xs[1]-m\*Xs[0])

g = modelo.support\_vectors\_[-1]

values\_y2 = m\*xx+(g[1]-m\*g[0])

Y=np.array(y)

X=np.array(x)

plt.scatter(X[:,0],X[:,1],c=Y)

plt.plot(xx,yy,'k-')

plt.scatter(vs[:,0],vs[:,1],s=80, facecolor = 'k')

plt.grid()

plt.show()

accu =accuracy\_score(y,yhat)

Prec = (precision\_score(y,yhat, average=None))

Reca = (recall\_score(y,yhat, average=None))

1. (4 puntos) Aplique un clasificador del tipo regresión logística y un SVM, en la base de datos “dataset\_4\_b.csv”. Indique cuál de los dos tiene un mejor resultado y porque creen que suceda esto. Justifique su respuesta con código, mediciones o figuras.

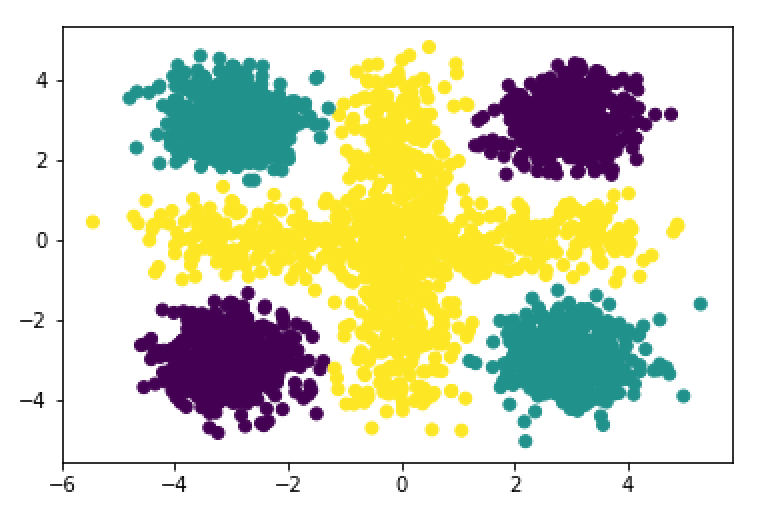
**Regresión Logistica: por medio de regresión logistica, el modelo arroja los siguientes resultados de Accuracy, Precision y Recall. Y se muestran que son bastante buenos, superando en todos los casos el 99%**

**https://lh4.googleusercontent.com/Otngeo_c4Eg02BFb-m9Z8chr6mPqcnQqBZ5zkcWAtgMwMDtiWnsQPCN46-o7cGFVjl1jaDoquCOSGx5PtawFjHaro0xQQIhwK5wfduN8dl-2TYI01XNOJT6fiudd2HlLFEVWCtqt**

**Modelo SVM: por medio del modelo SVM, se obtiene los siguientes resultados de Accuracy, Precision y Recall.**

**https://lh6.googleusercontent.com/qkHg3qpQZZQ0CBAJsOk0W5yaYuq2KWq63hTiqNFAA_hr9_N3RoMT2P4e-y00PZiIBb7UD3duGCHwSHUn7Ant-mtJBGCZBTtWeN15NJO2kawGXEOec8ImsIy0Ew_-G6dJ1747_9VU**

**Se ve que en este caso los resultados son bastante superiores de medición, aunque similares ya que superan también el 99%.**

****

**En regresión lineal se ve que se comporta como un polinomio de 2do orden. También se comprueba con el SVM, dando los resultados de Y nuevos. En ambas ocasiones los resultados de Accuracy, Precision, Recall, son próximos al 99% y muy buenos. En este caso el SVM tuvo mejores mediciones que la RL. La gráfica pasada permite visualizar como quedan ordenados los datos mediante los métodos, una manera muy definida.**