

Parte II

RECONOCIMIENTO DE FORMAS

Clustering y aprendizaje
supervisado

Índice de contenido

1.Algoritmos de clustering.....	3
1.1 K-Medias.....	4
1.1.1 Marco teórico.....	4
1.1.2 Resultados.....	4
1.2 Secuencial.....	49
1.2.1 Marco teórico.....	49
1.2.2 Resultados.....	49
2.Aprendizaje supervisado.....	62
2.1 Perceptrón.....	62
2.1.1 Marco teórico.....	62
2.1.1 Resultados.....	63
2.2 Ensemble.....	70
2.2.1 Marco teórico.....	70
2.2.2 Resultados.....	70
2.3 Reglas Granuladas.....	77
2.3.1 Marco teórico.....	77
2.3.2 Resultados.....	77

1. Algoritmos de clustering

Los algoritmos de clustering son técnicas designadas como aprendizaje no supervisado, al disponer originalmente de un data set de datos no etiquetados en los cuales es necesario establecer cuáles son los distintos grupos naturales.

1.1 K-Medias

1.1.1 Marco teórico

El algoritmo de las K-Medias consiste en que, a partir de unos clusteres iniciales -elegidos al azar (opción 1) o de manera óptima (opción 2, escogiendo como clusters iniciales aquellos que más alejados entre sí dentro de los posibles clusters que se pueden crear con el data set)- cuyos centroides serán la media de sus componentes, de forma iterativa estos se irán reorganizando añadiendo los datos a aquel cuya distancia al dato es menor. Se recalcularán los centroides y se repetirá el proceso hasta que no haya variaciones en los clusters.

1.1.2 Resultados

Se ha realizado el estudio del índice de compactación, el índice de Dunn y el índice de Fisher para todos los data sets de la asignatura para los distintos valores del parámetro crítico K, primero mediante inicialización aleatoria y después también mediante inicialización óptima.

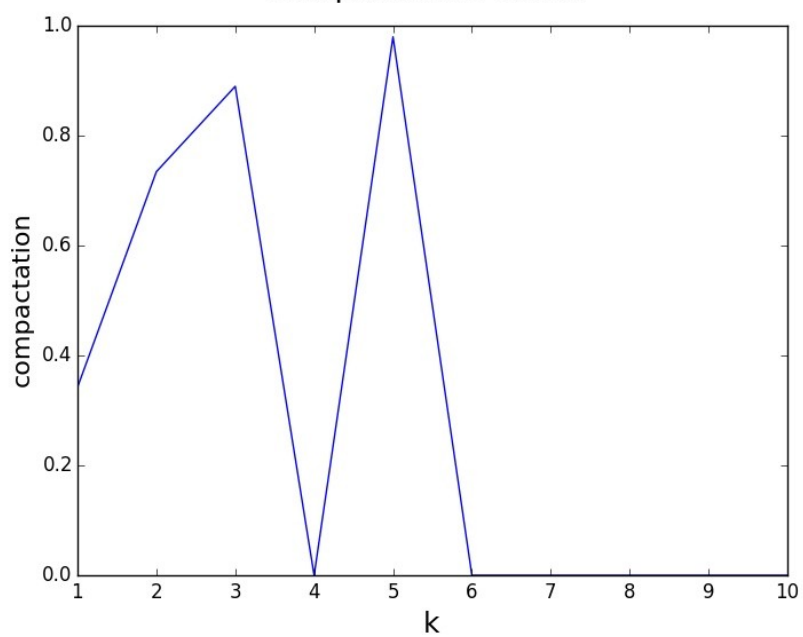
Se ha probado en la mayoría de data sets dos rangos de K: entre 1 y 10 (donde se producían la mayoría de grandes diferencias), y entre 1 y 100. En algunos casos, debido a el tamaño del data set, probar k entre 1 y 100 suponía demasiado tiempo de ejecución por lo que se escogió el rango 1:30.

a) Iris

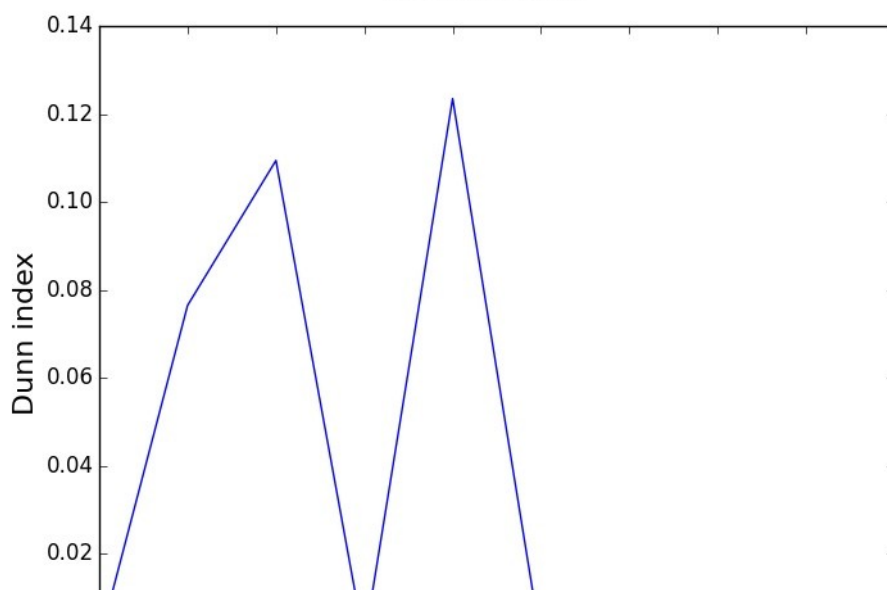
- Inicialización aleatoria

-K 1:10

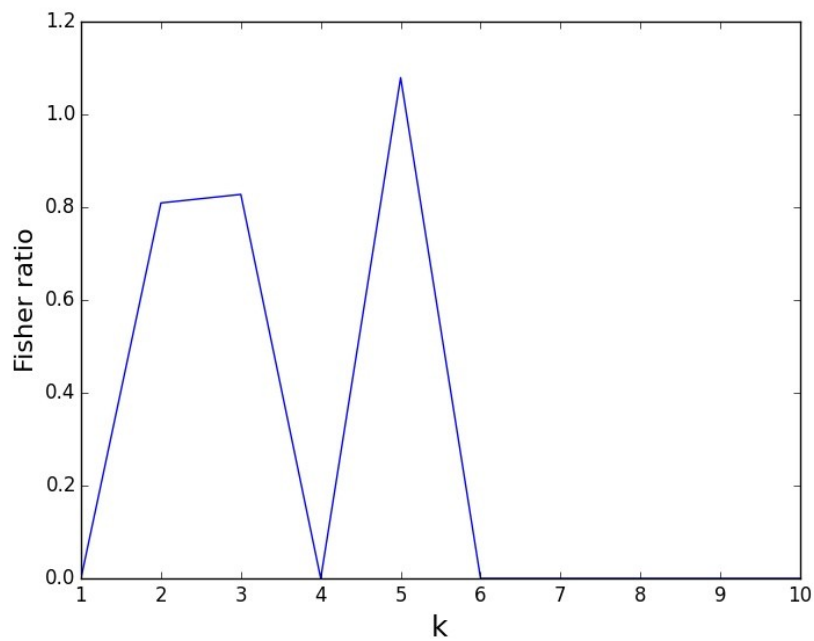
Compactation Index



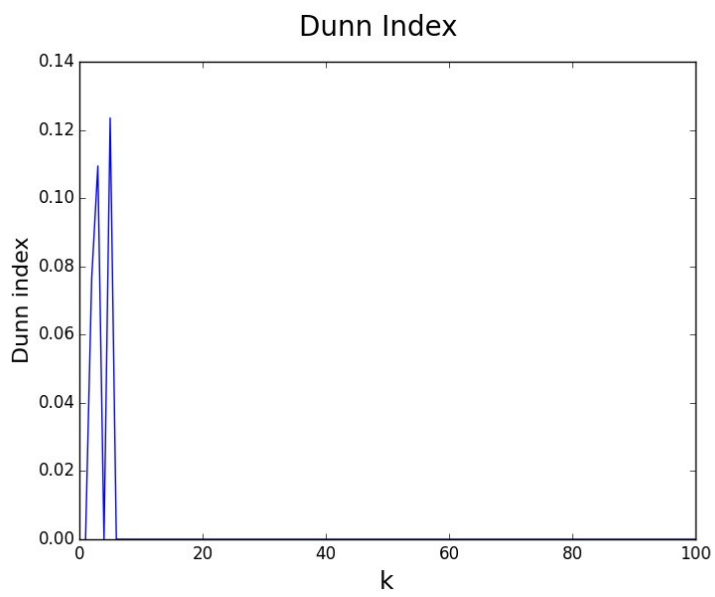
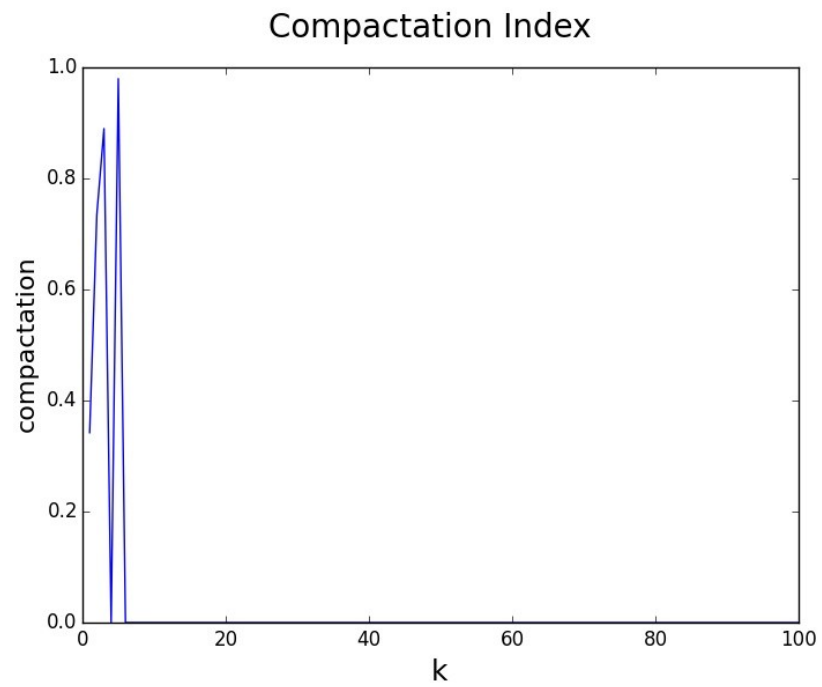
Dunn Index

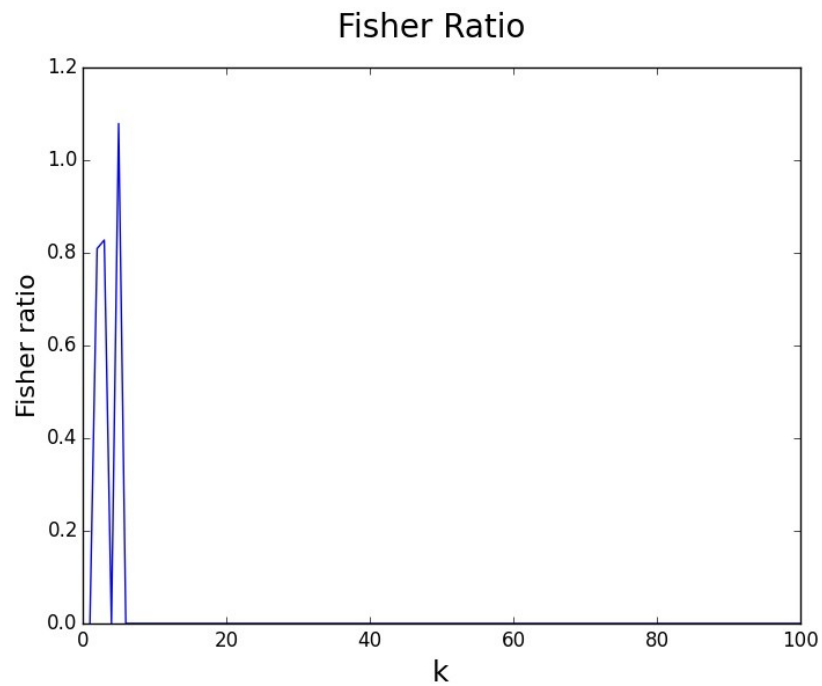


Fisher Ratio



-K 1:100

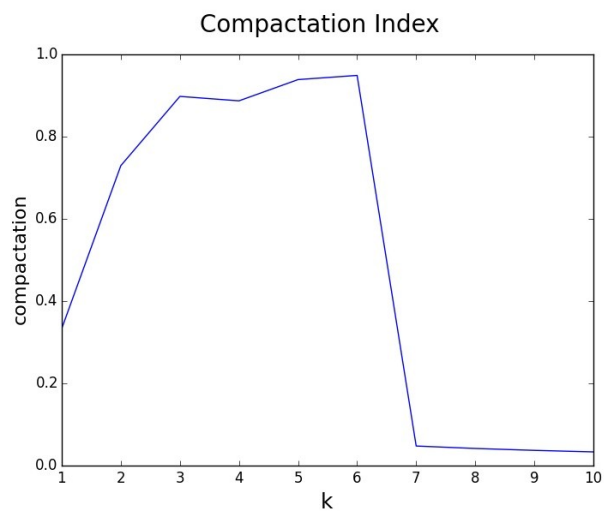


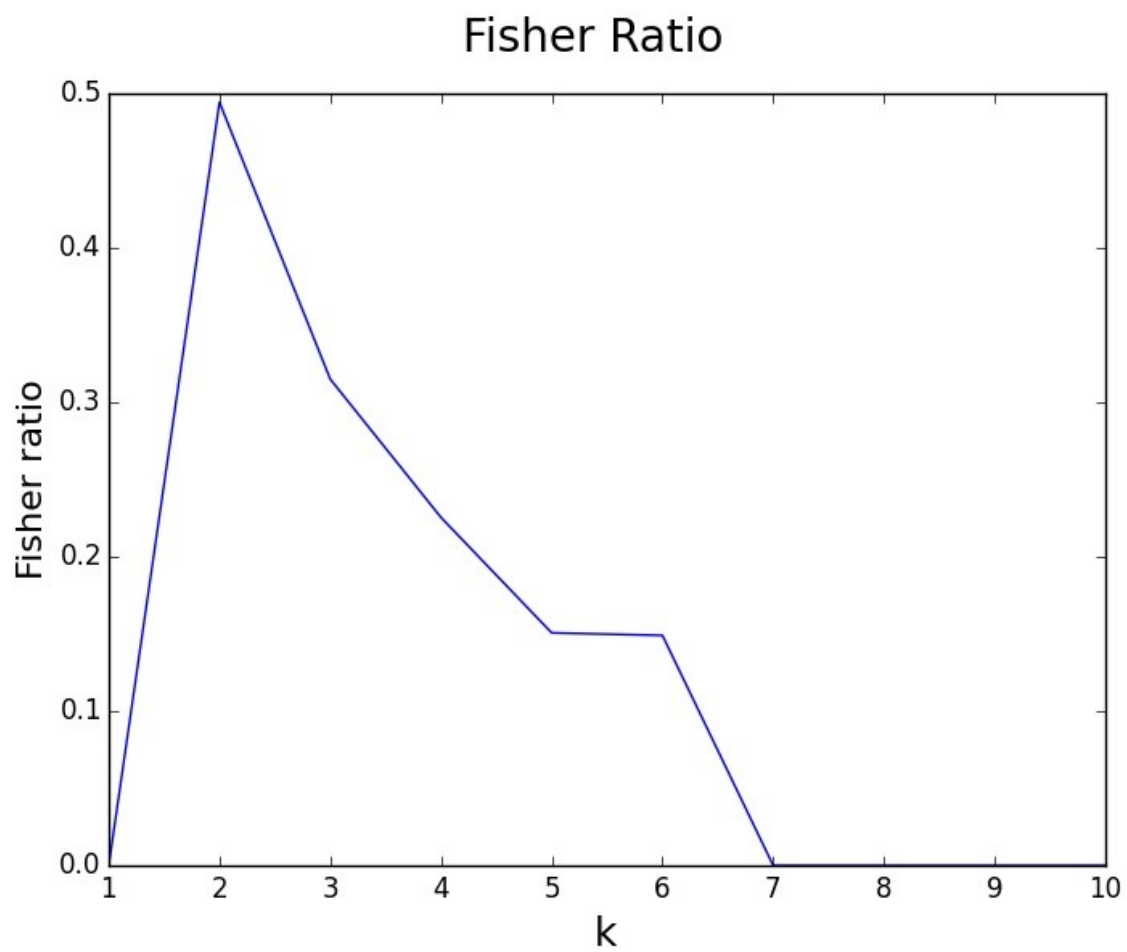
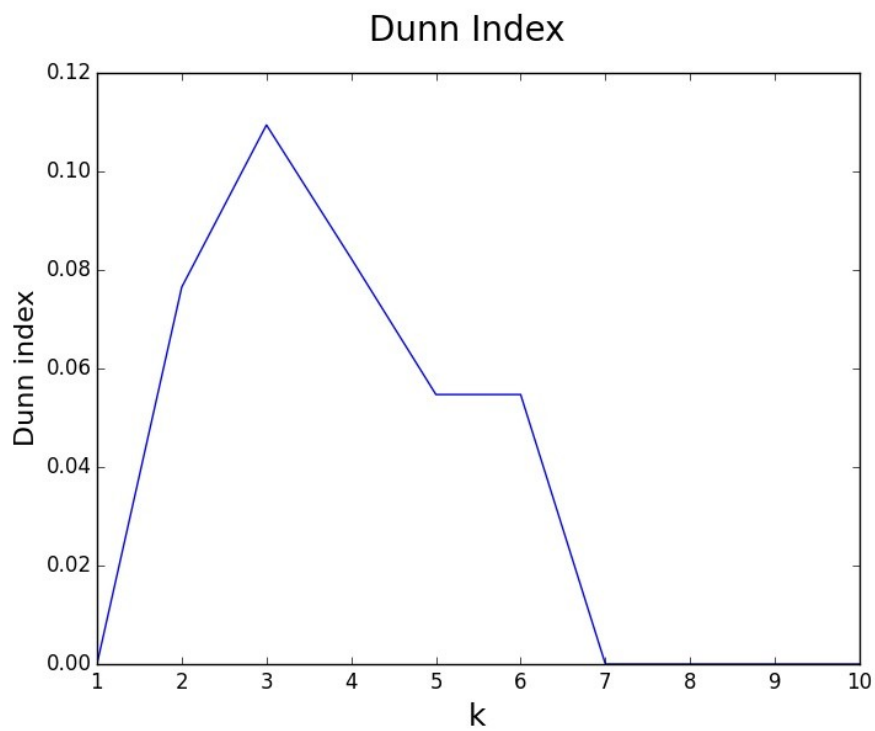


Vemos que en casi todas las gráficas hay dos picos en los que los índices de calidad se maximizan: para $k=3$ (número real de clusters) y para $k=5$ por lo que optimizaríamos k en uno de estos valores.

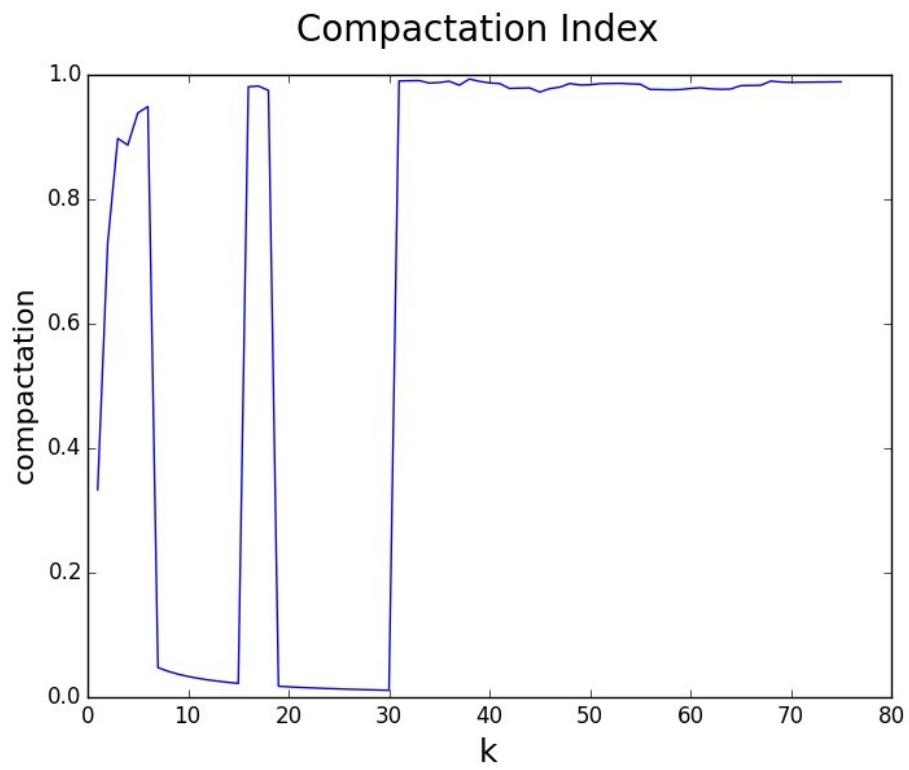
- Inicialización óptima

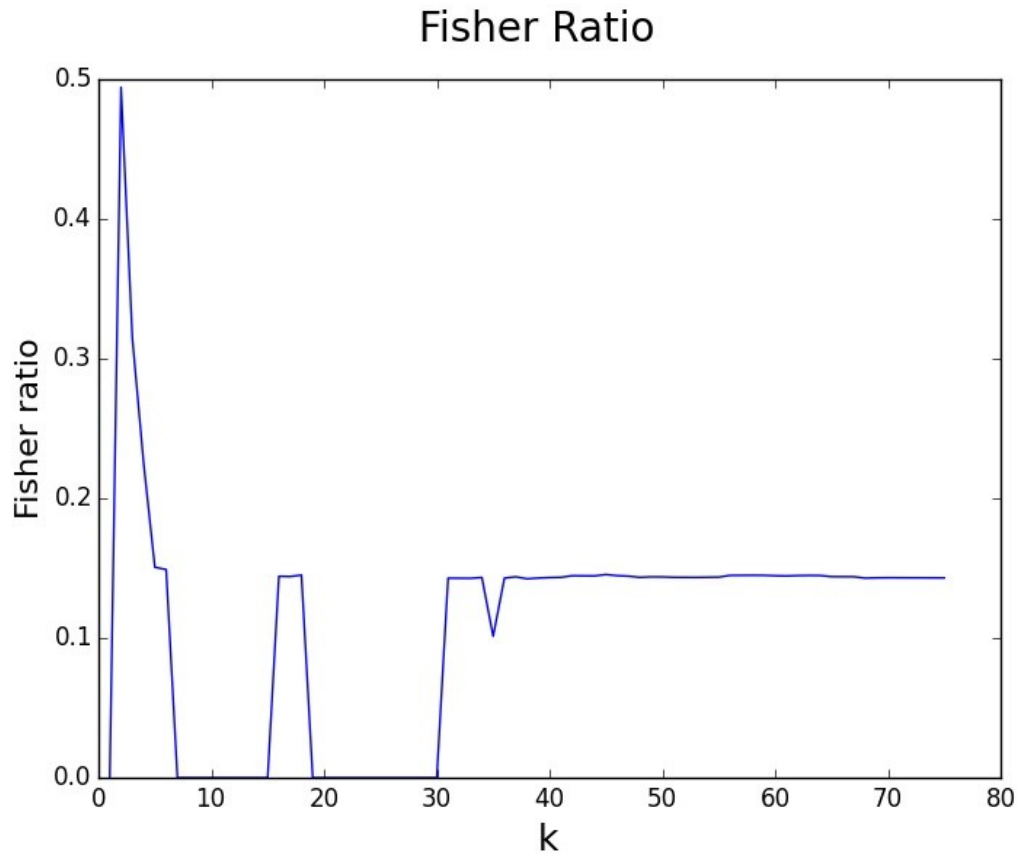
-K 1:10





-K 1:100





De nuevo los valores máximos se han alcanzado para $k=3$ (en algunos casos $k=2$: como vimos en la parte 1 de la asignatura, hay dos clases que están muy pegadas en este data set y una más separada).

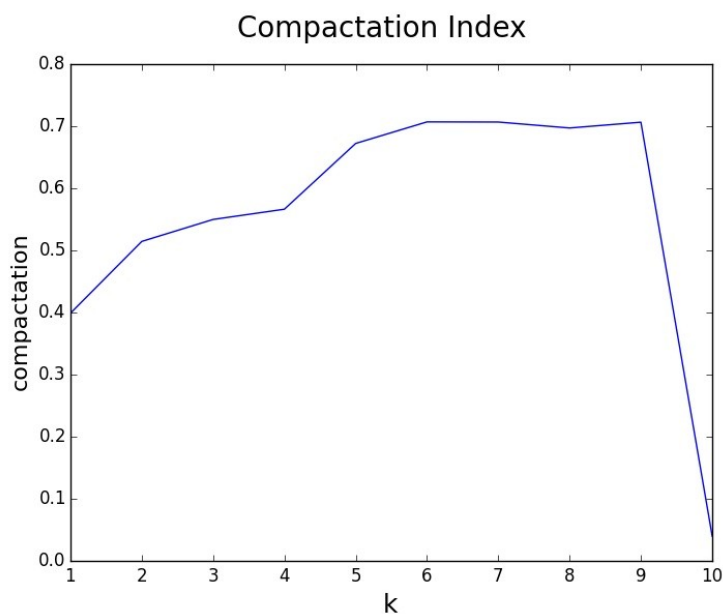
Para valores grandes de k , los índices se descontrolan, debido al pequeñísimo tamaño de los clusters.

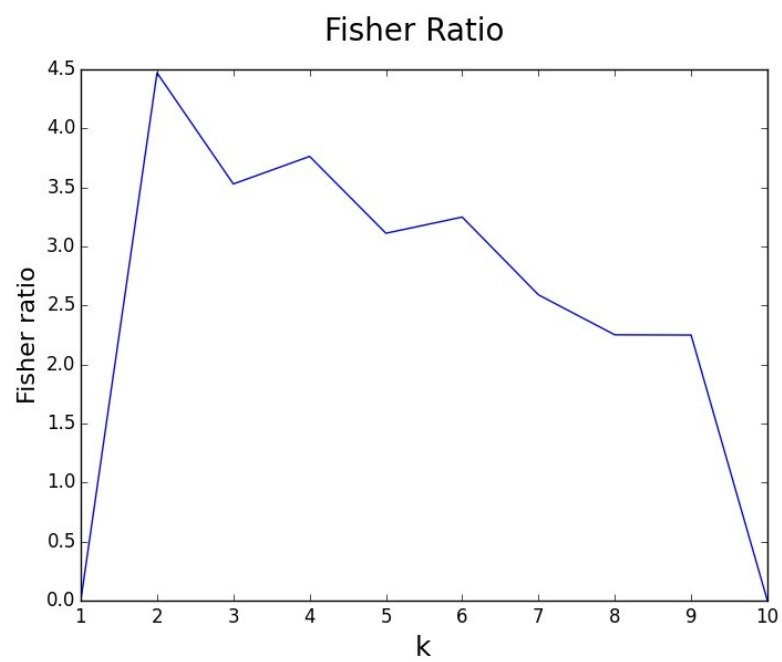
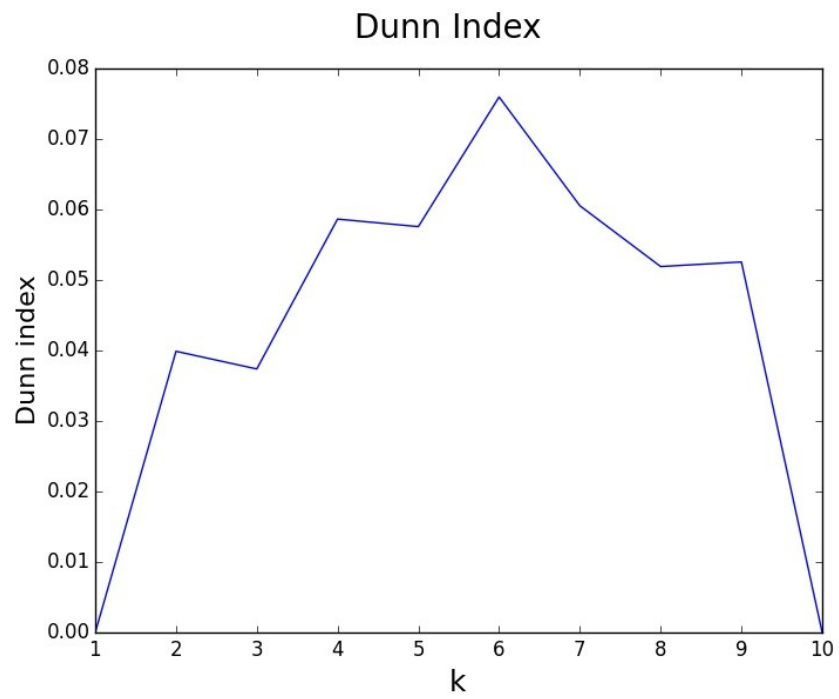
La inicialización óptima parece que provoca que llegue antes a valores altos de los índices.

b)Wine

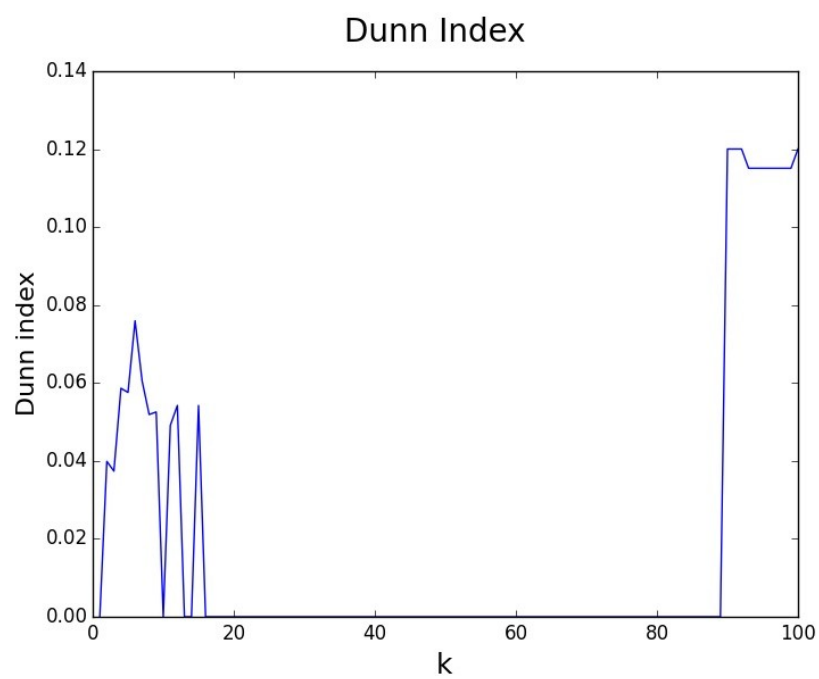
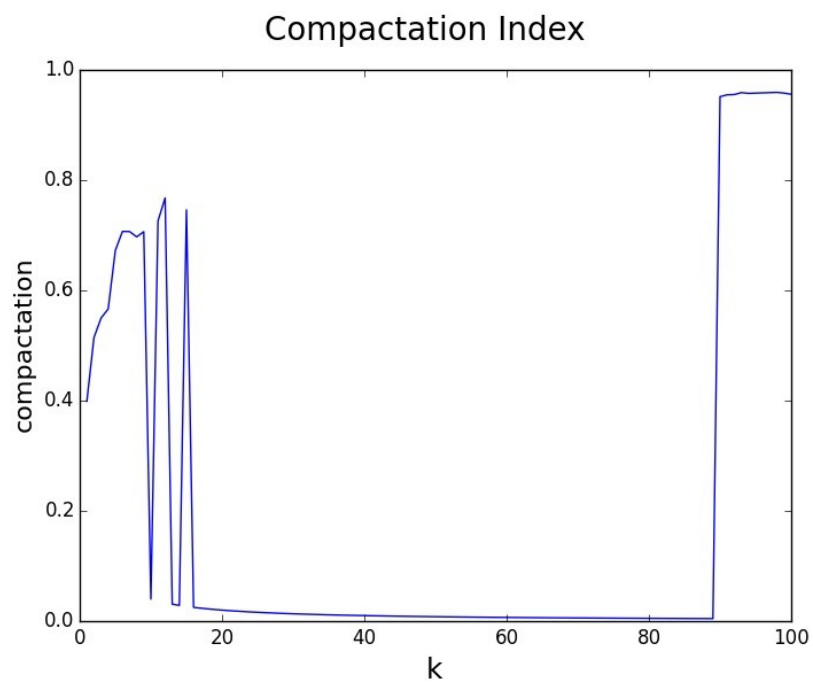
- Inicialización aleatoria

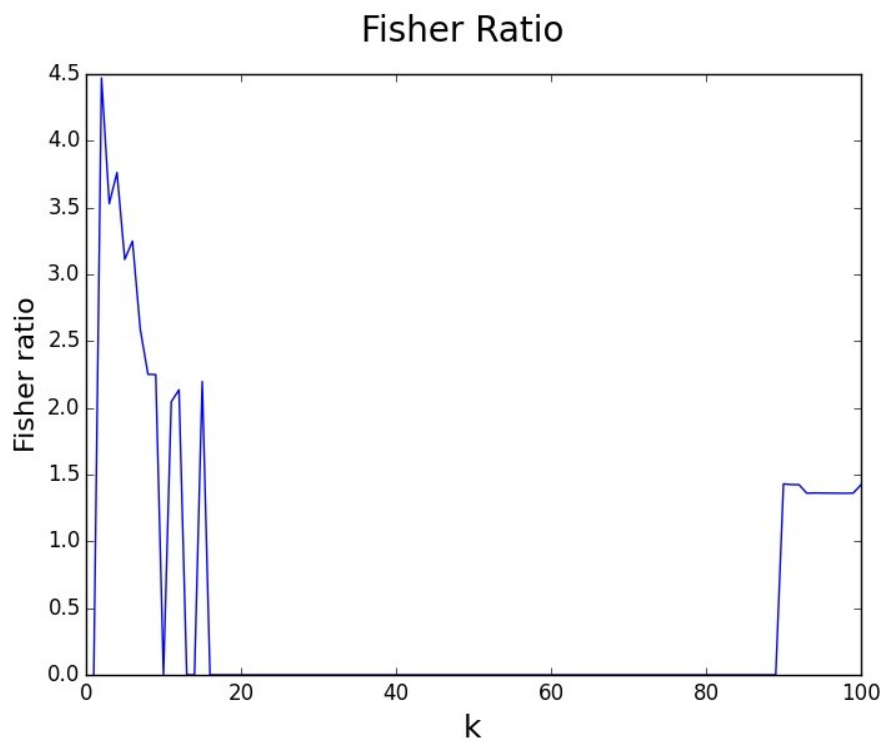
-K 1:10





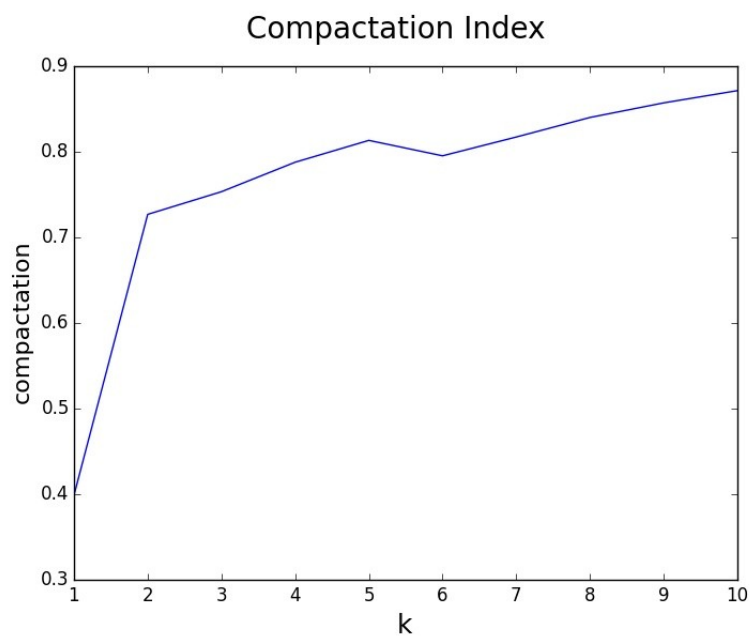
-K 1:100

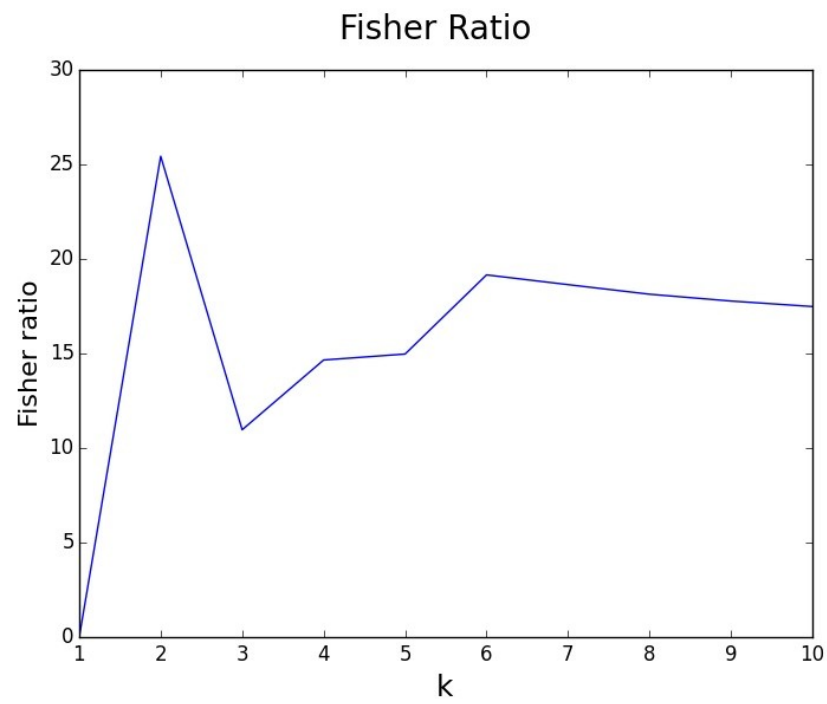
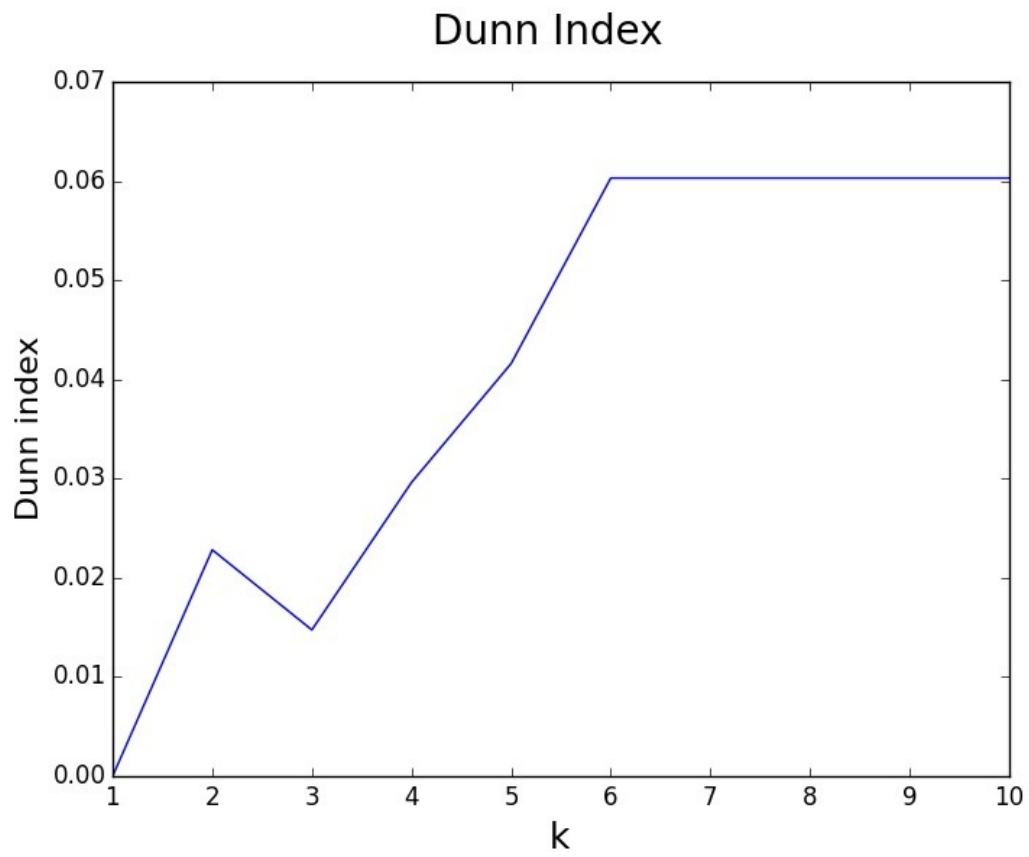




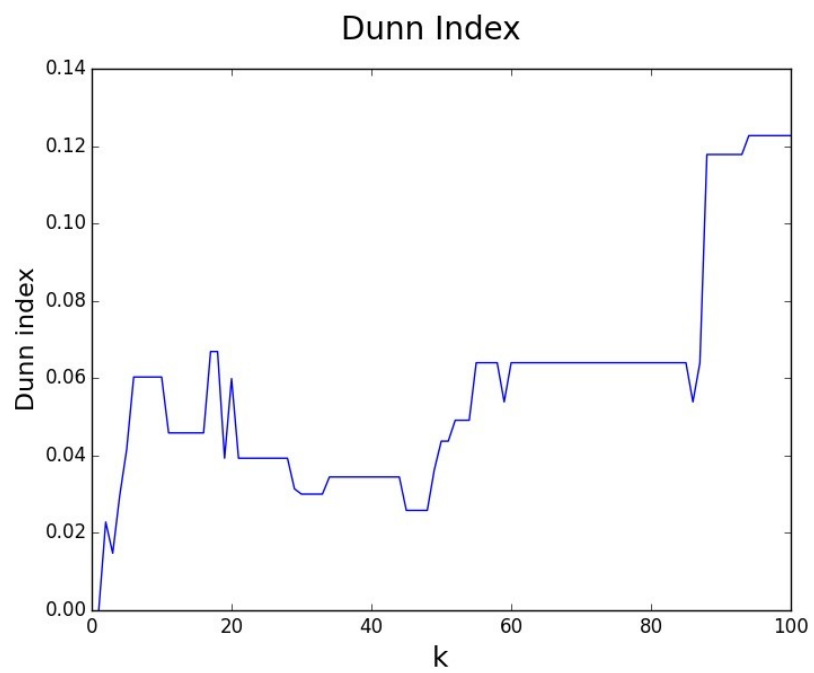
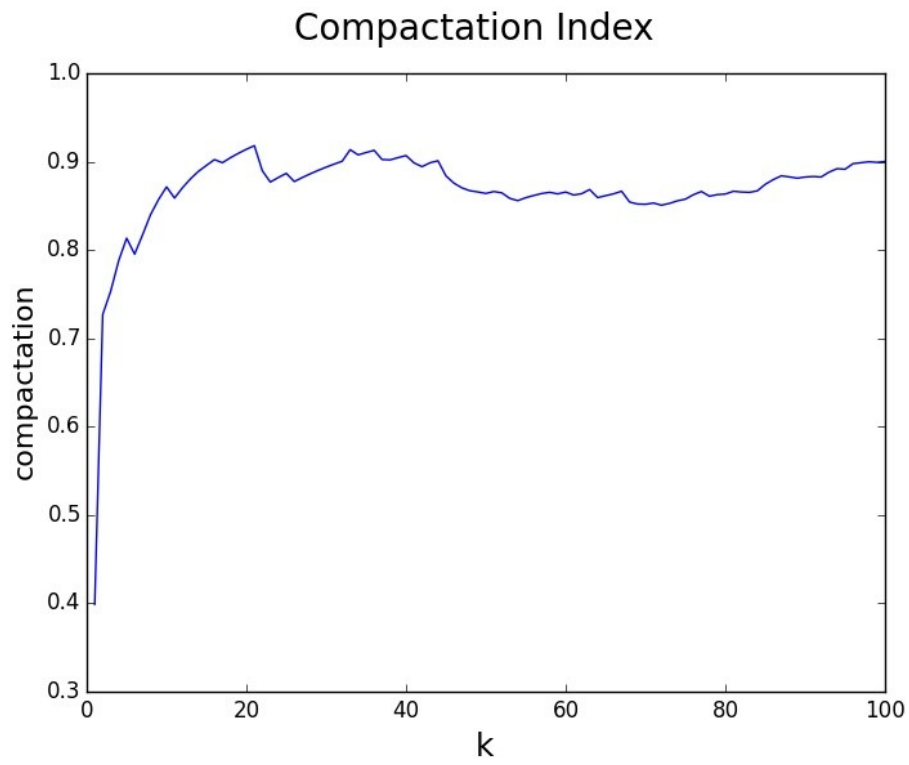
- Inicialización óptima

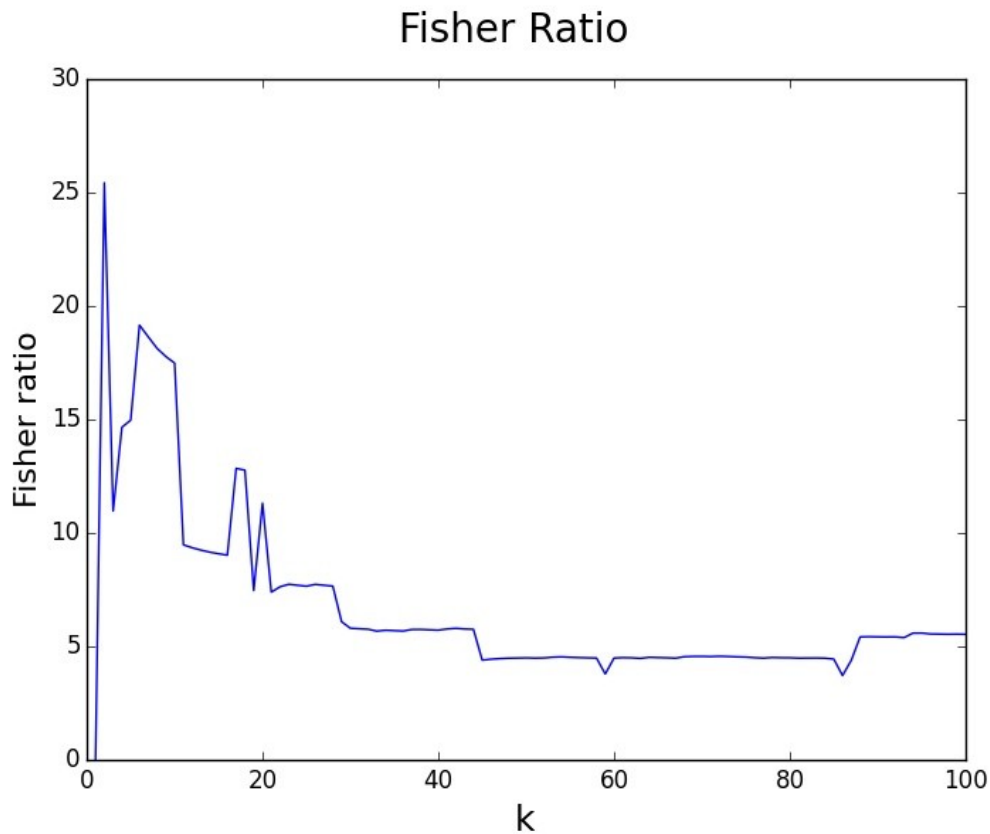
-K 1:10





-K 1:100



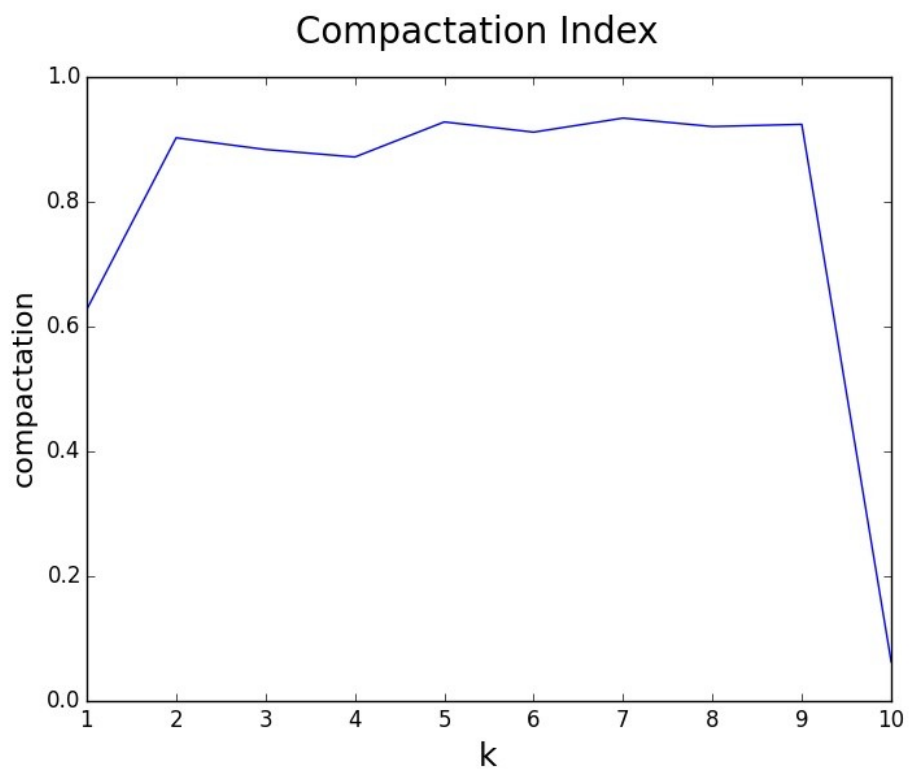


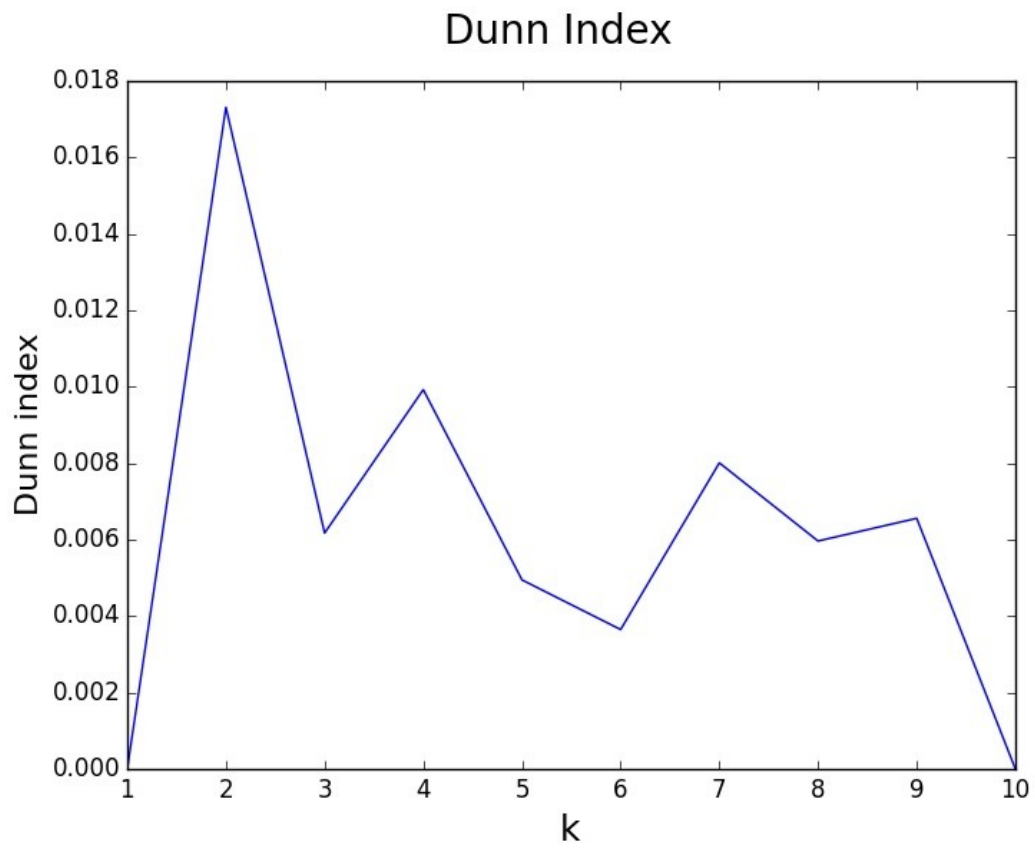
A la vista de los resultados, vemos que los índices se sitúan en valores altos a partir de $k=2,3$ (coincidiendo con el número real de clases) para ambas inicializaciones (crecen muy rápido ya para valores muy pequeños de k), manteniéndose constantes o descontrolándose para valores muy altos lo que, si no conociéramos el número total de clases, evidenciaría que este último no es demasiado alto.

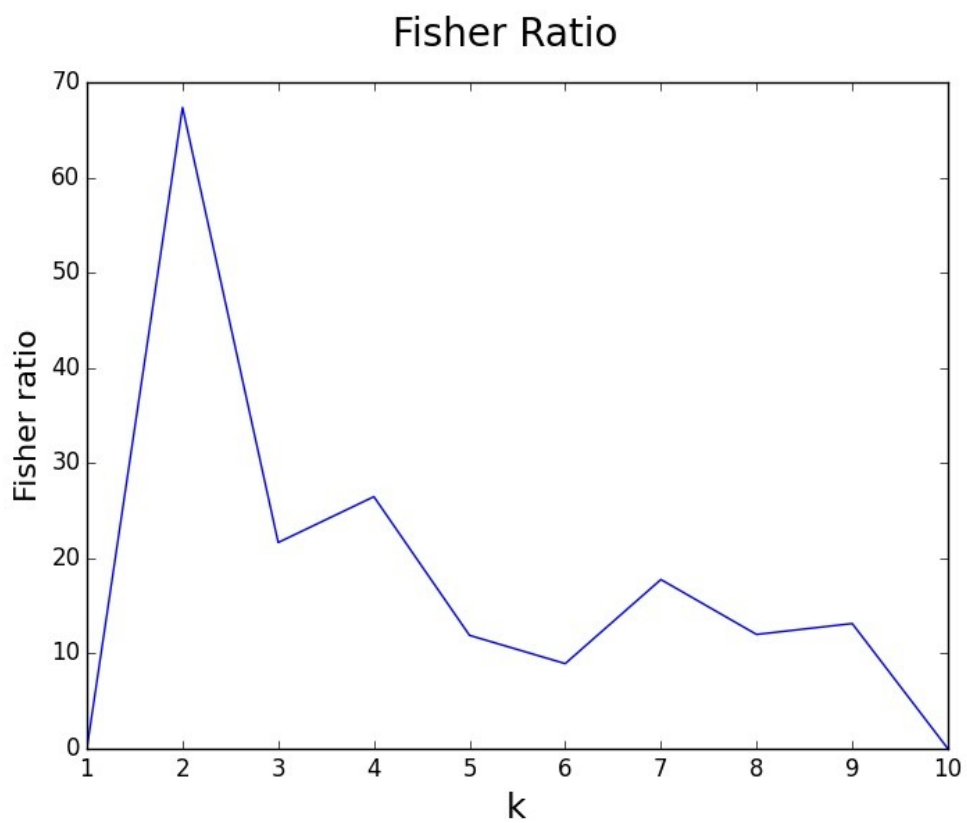
c) Cáncer

- Inicialización aleatoria

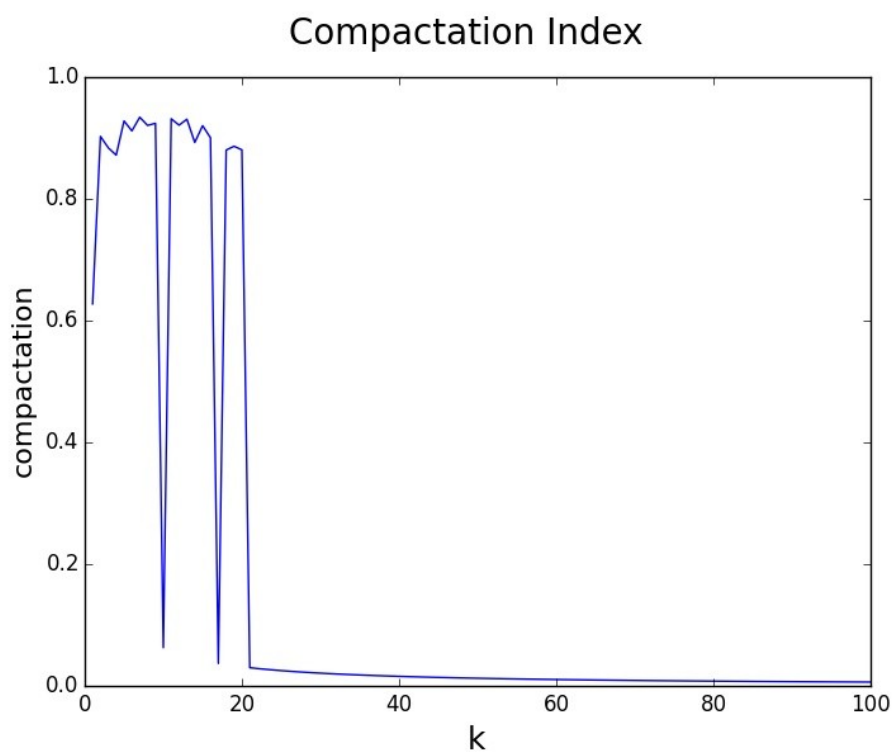
-K 1:10

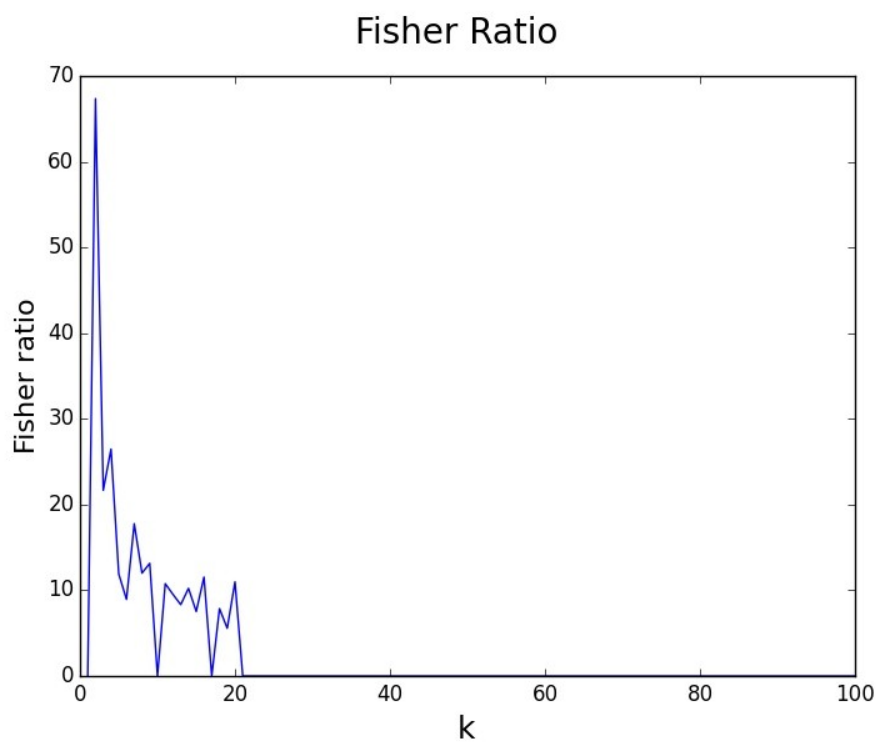
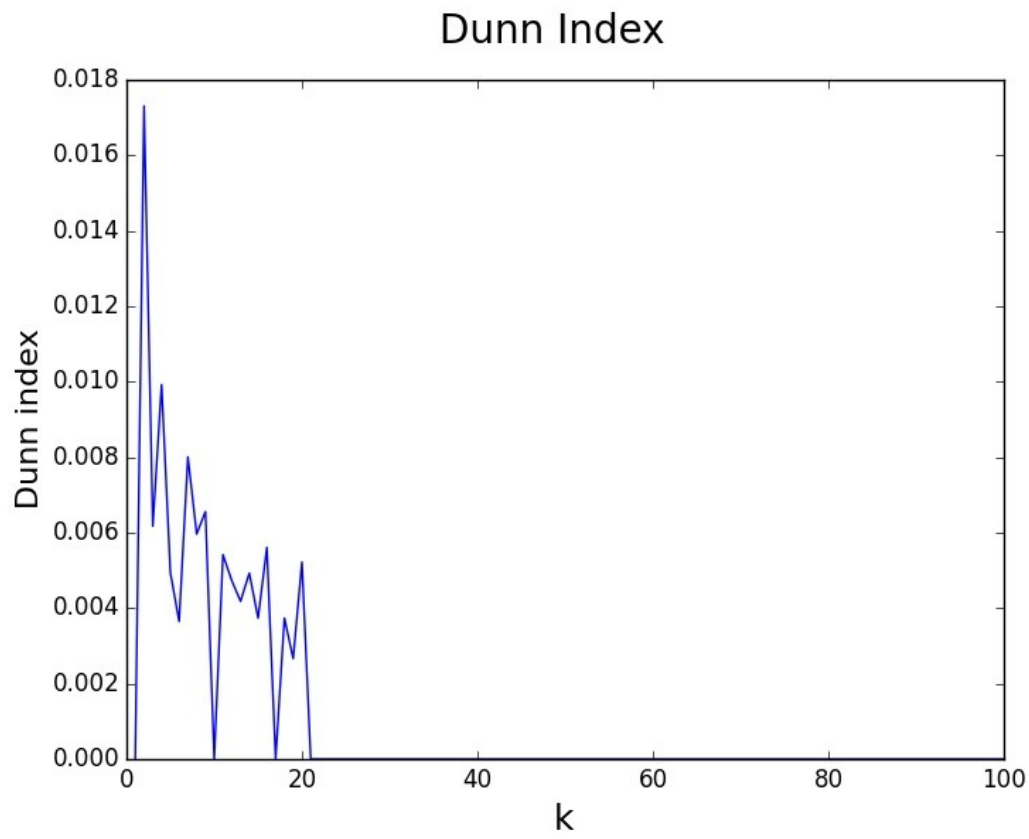






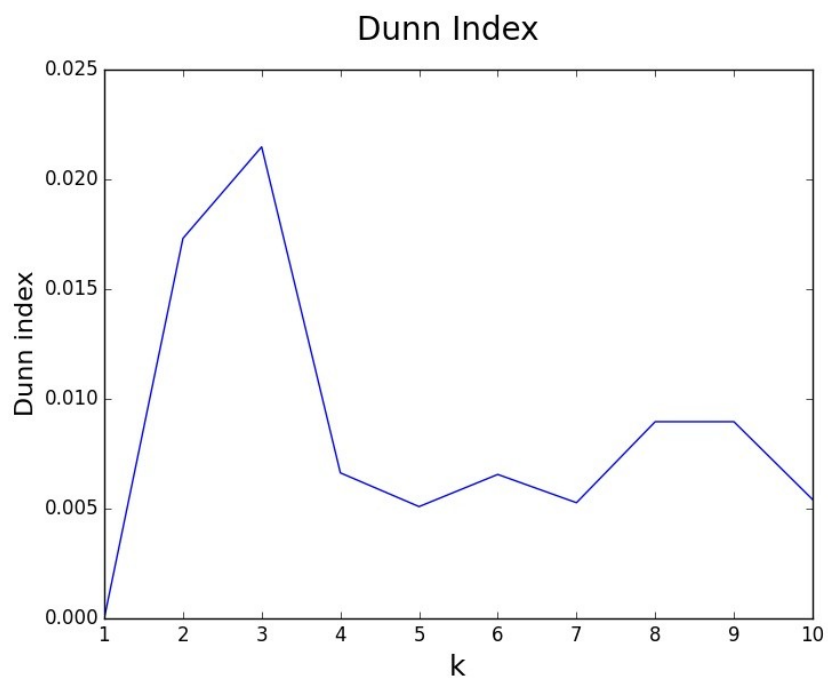
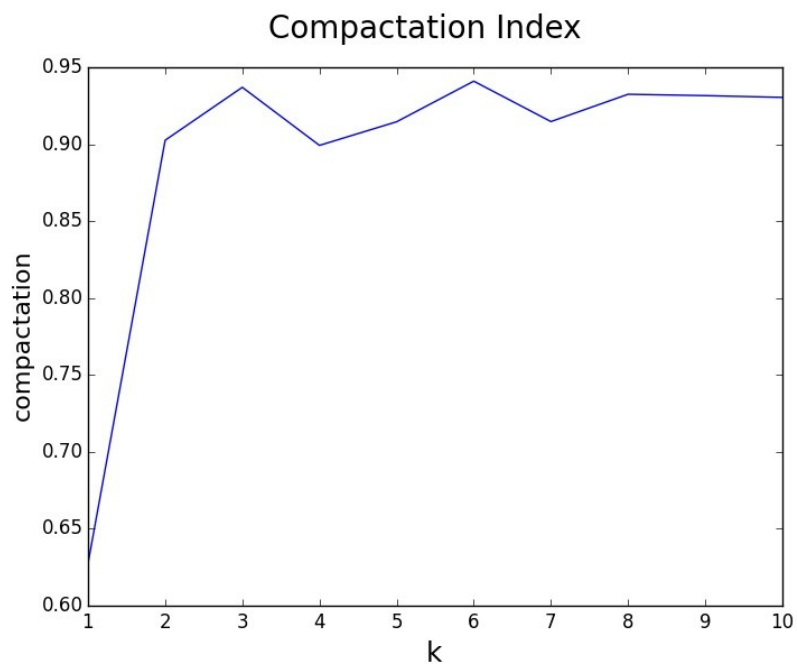
-K 1:100

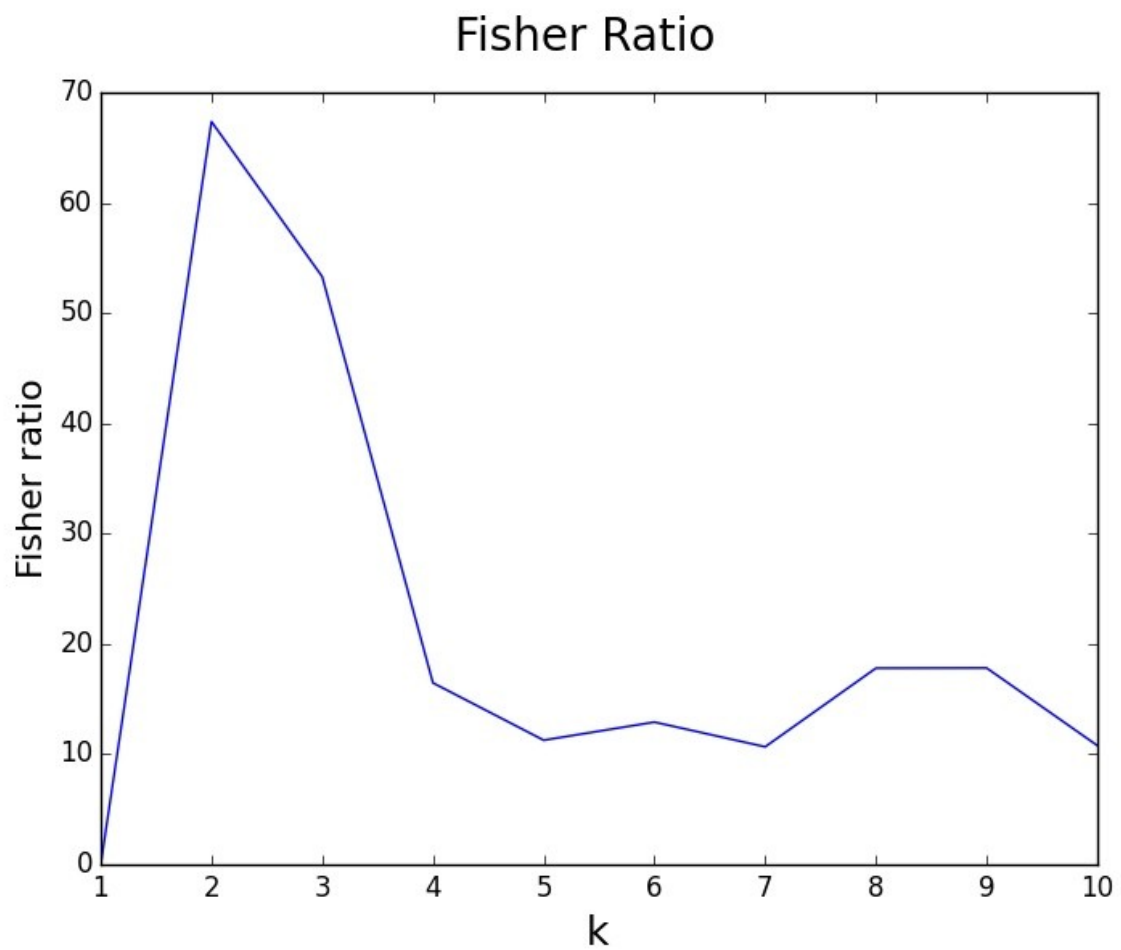




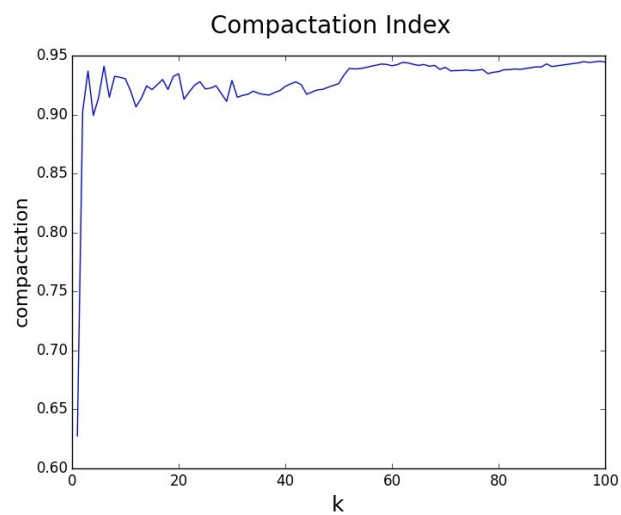
- Inicialización óptima

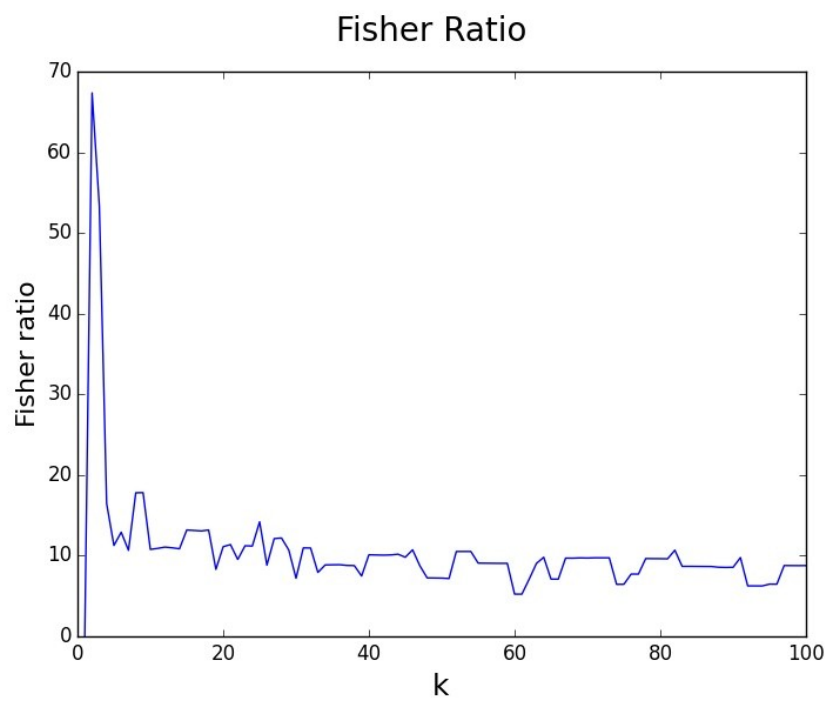
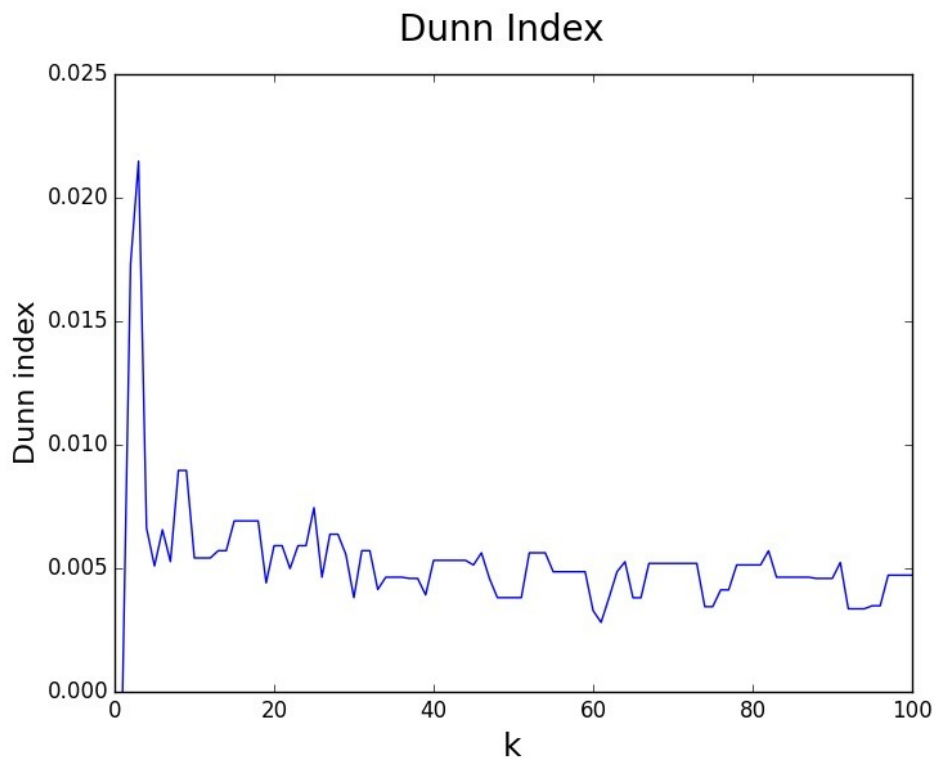
-K 1:10





- K 1:100



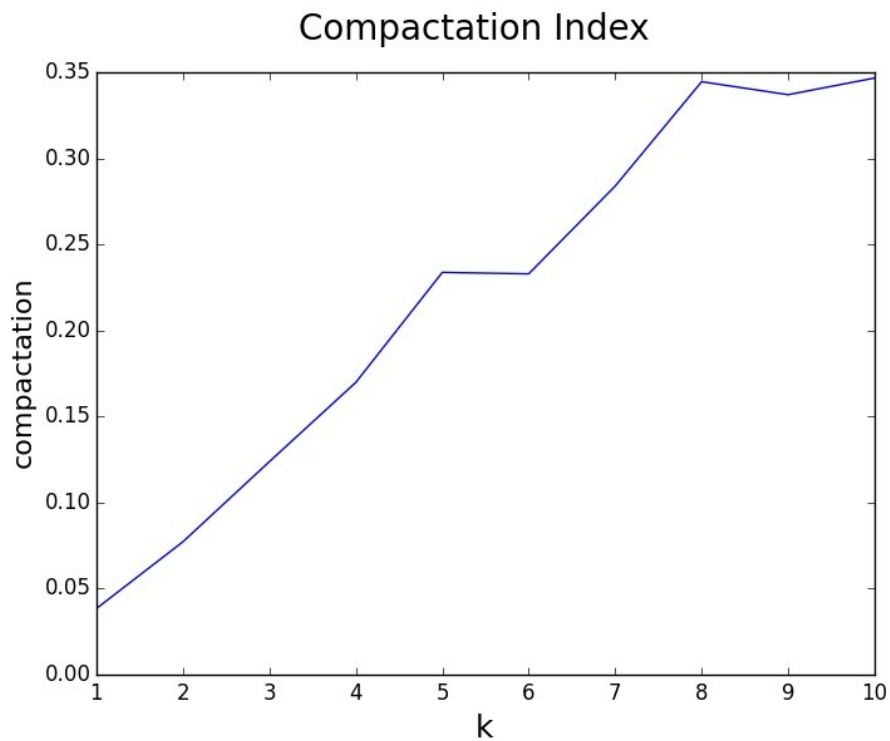


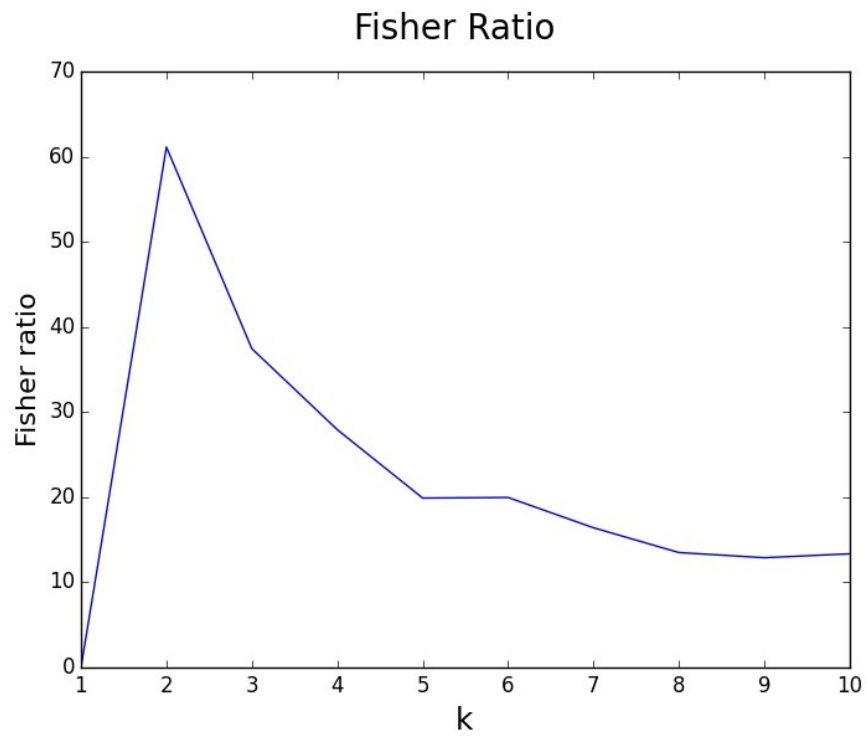
Para este data set los resultados del K-medias son muy claros: en casi todas las gráficas, el máximo de los índices de calidad se alcanza para $k=2$ que coincide con el número de clases reales (B y M), lo que confirma la credibilidad del algoritmo.

d) Isolet

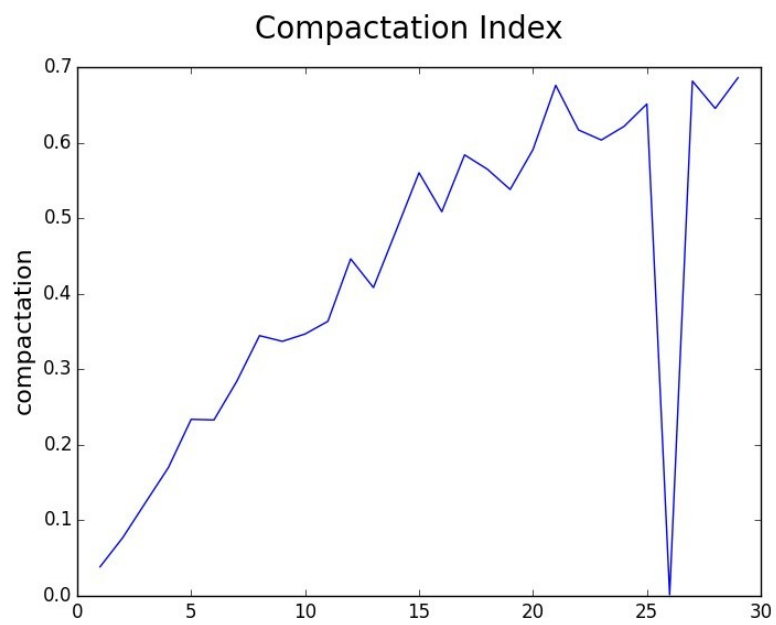
- Inicialización aleatoria

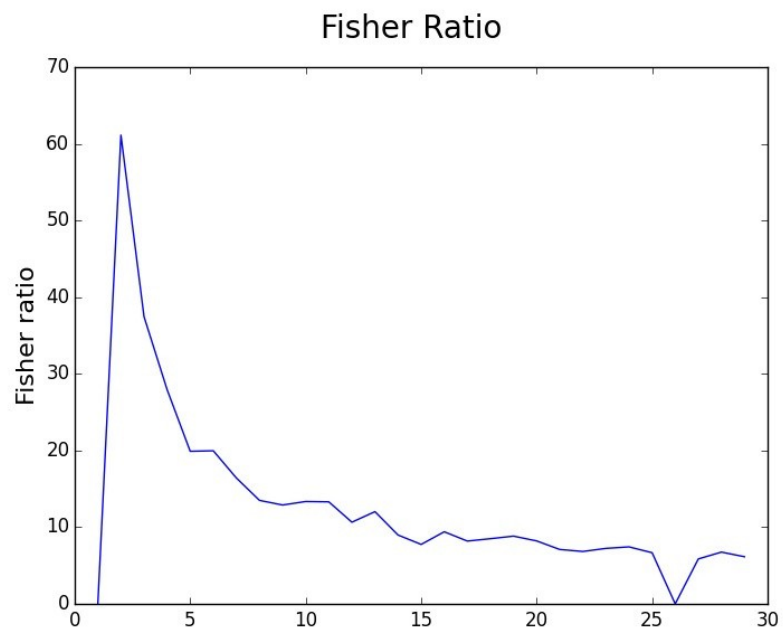
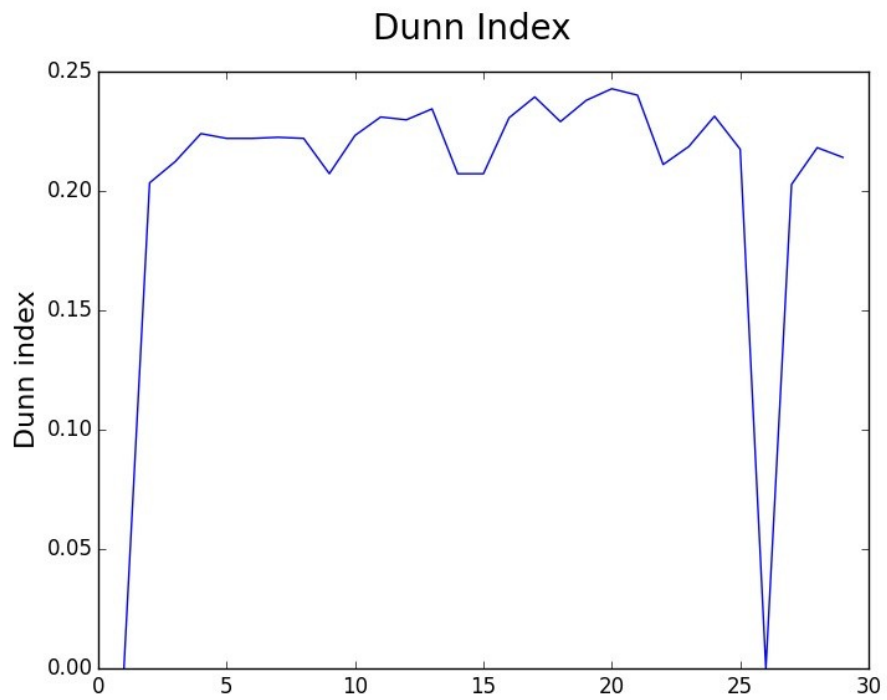
-K 1:10





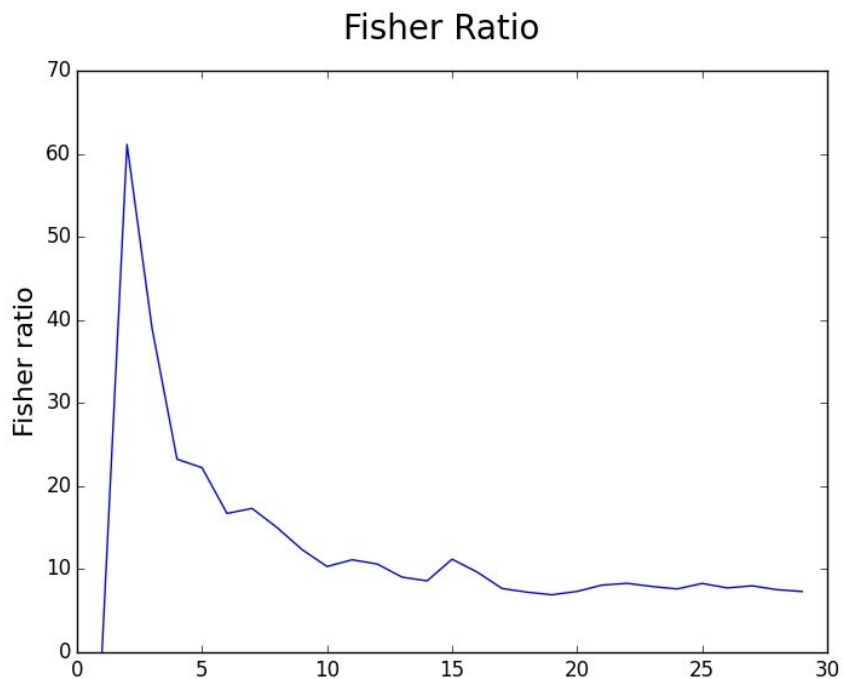
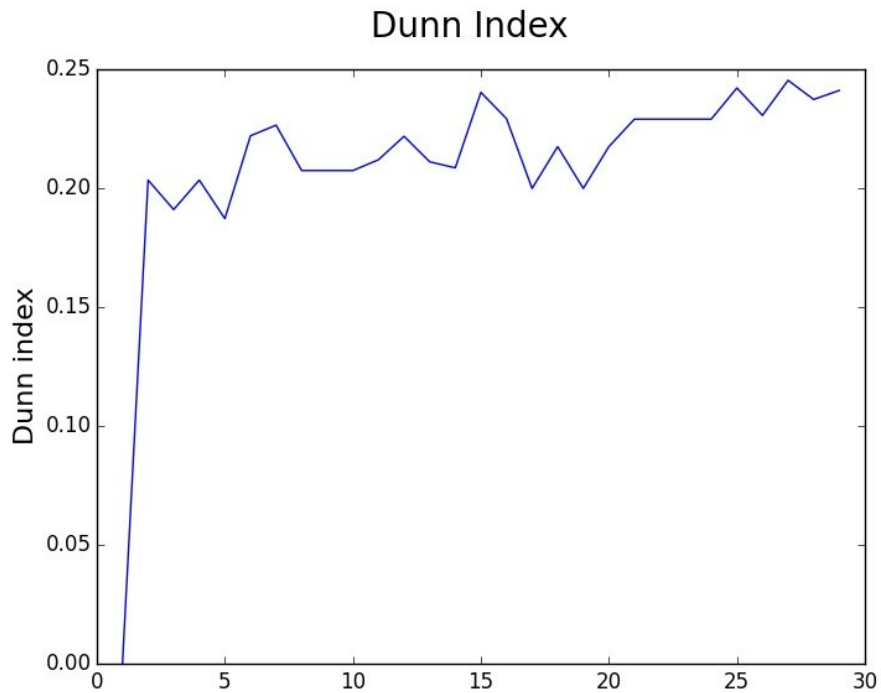
-K 1:30





- Inicialización óptima

-K 1:30



Los resultados para este data set han resultado algo confusos. En primer lugar, el comportamiento del algoritmo con inicialización aleatoria e inicialización óptima es claramente diferenciado. Los índices de calidad de la inicialización aleatoria crecen progresivamente hasta alcanzar un máximo en la zona donde k toma el valor del número de clases reales (26) y justo después realiza una bajada

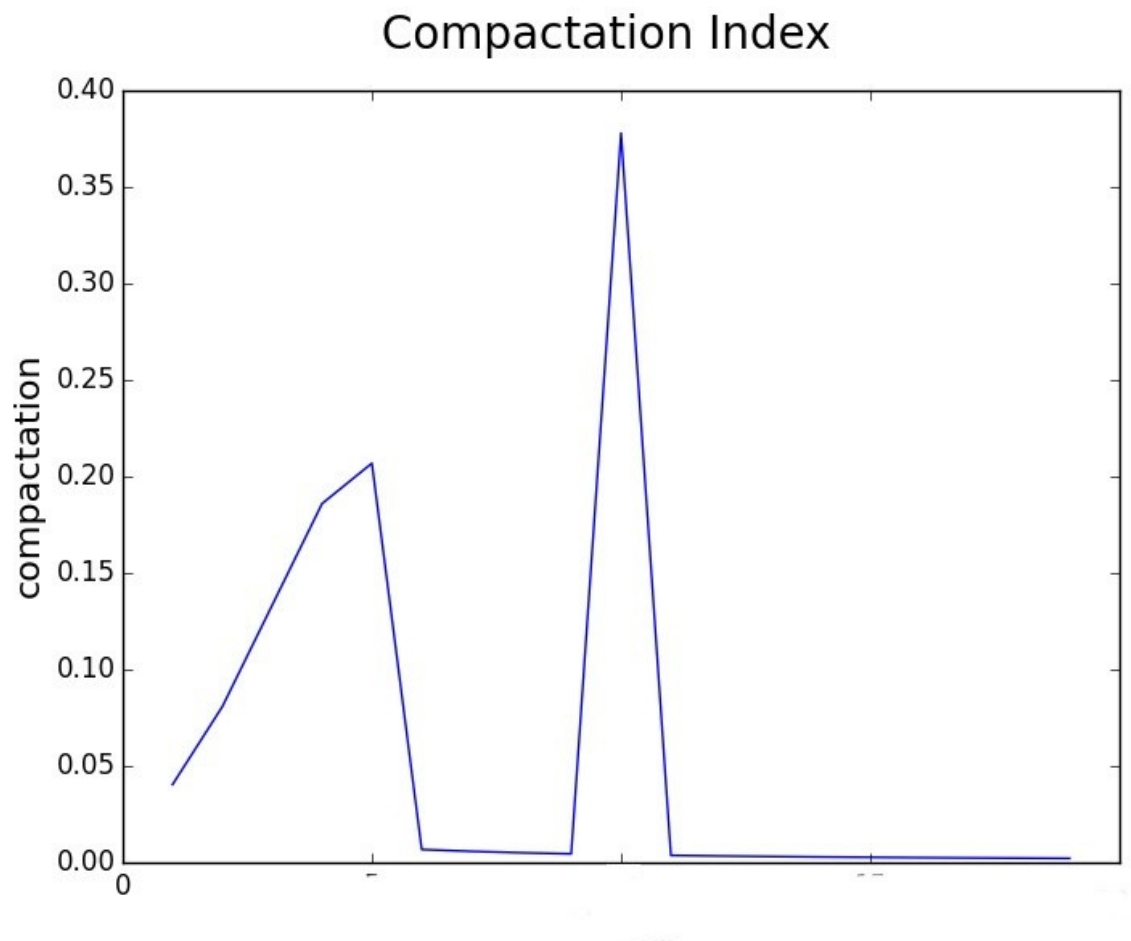
brusca. Esto podría explicarse porque a partir de ese número, se comienzan a crear clases innecesarias resultando en una distribución nada compacta.

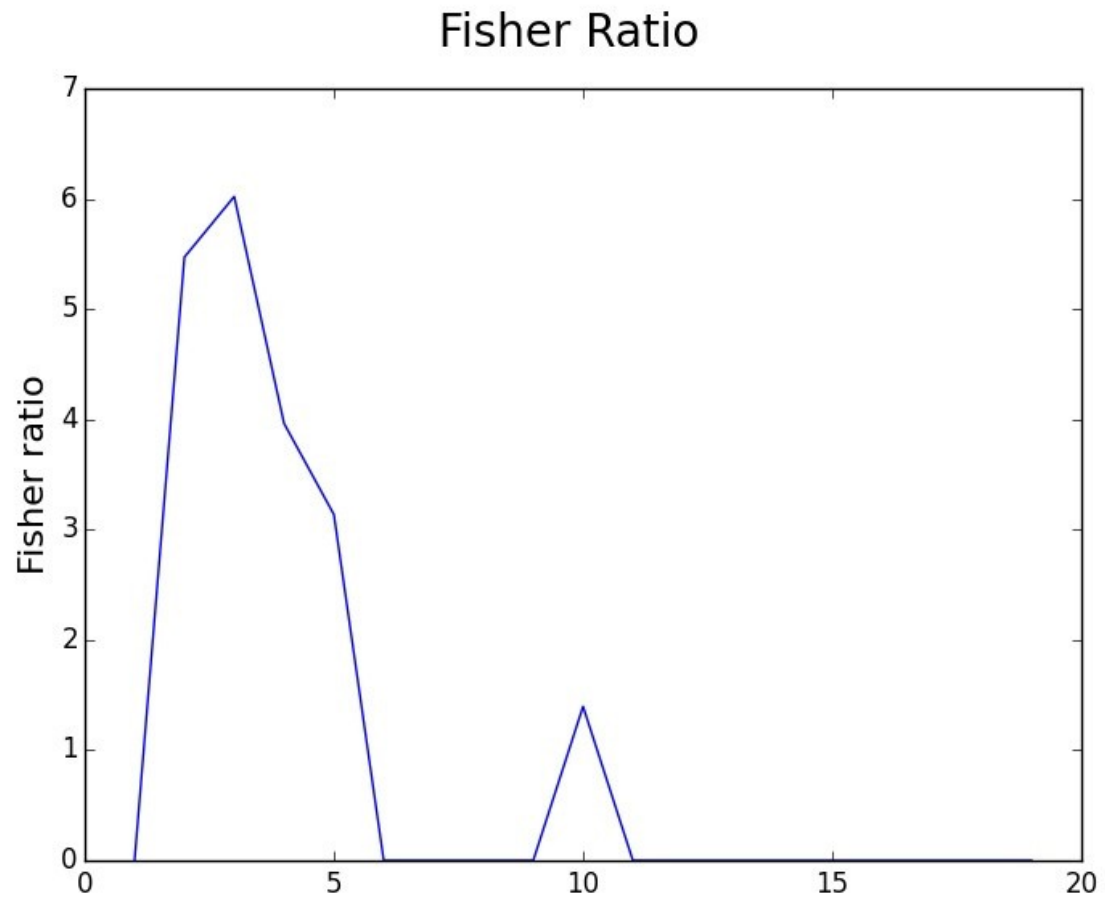
Con la inicialización óptima, se cumple como en casos anteriores, que los índices de calidad comienzan a crecer muy pronto (para valores de k bajos) y se mantienen bastante regulares. Llama la atención el índice de Fisher que alcanza en ambos casos un claro máximo para $k=5$. Tal vez, se puedan segmentar los datos en 5 grupos de clases bastante separados entre sí.

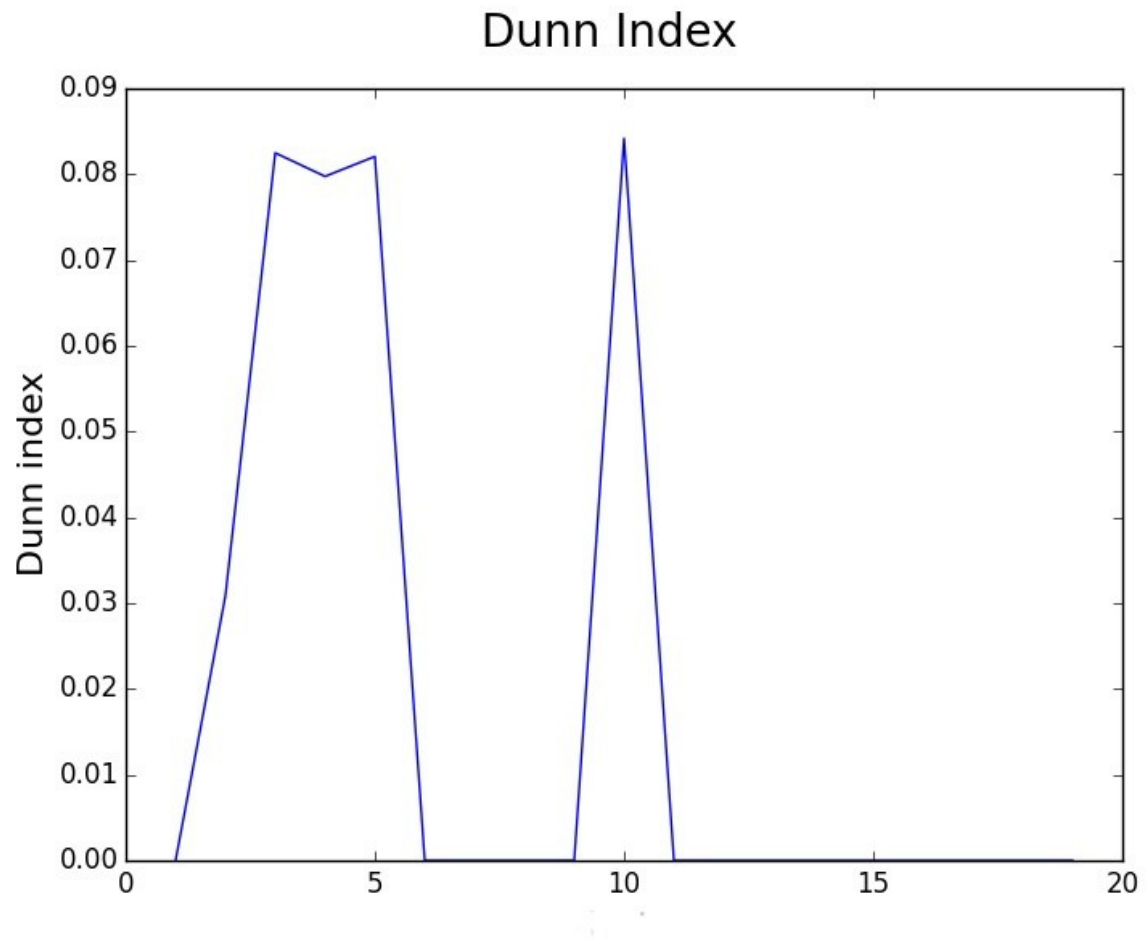
e) MNIST

- Inicialización aleatoria

-K 1:20



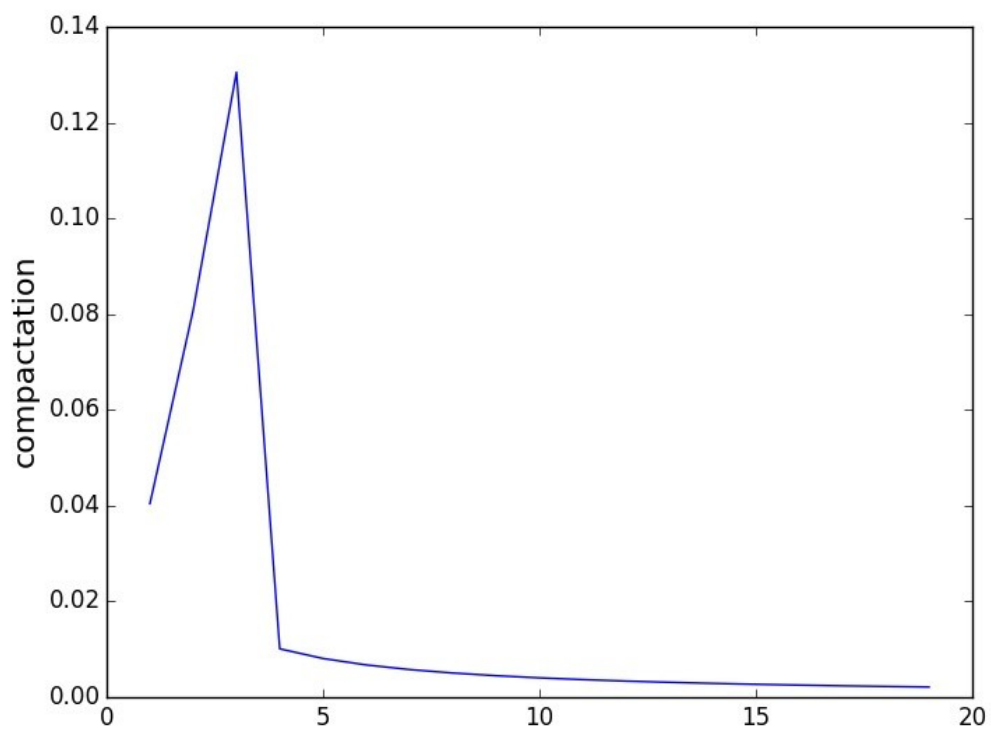




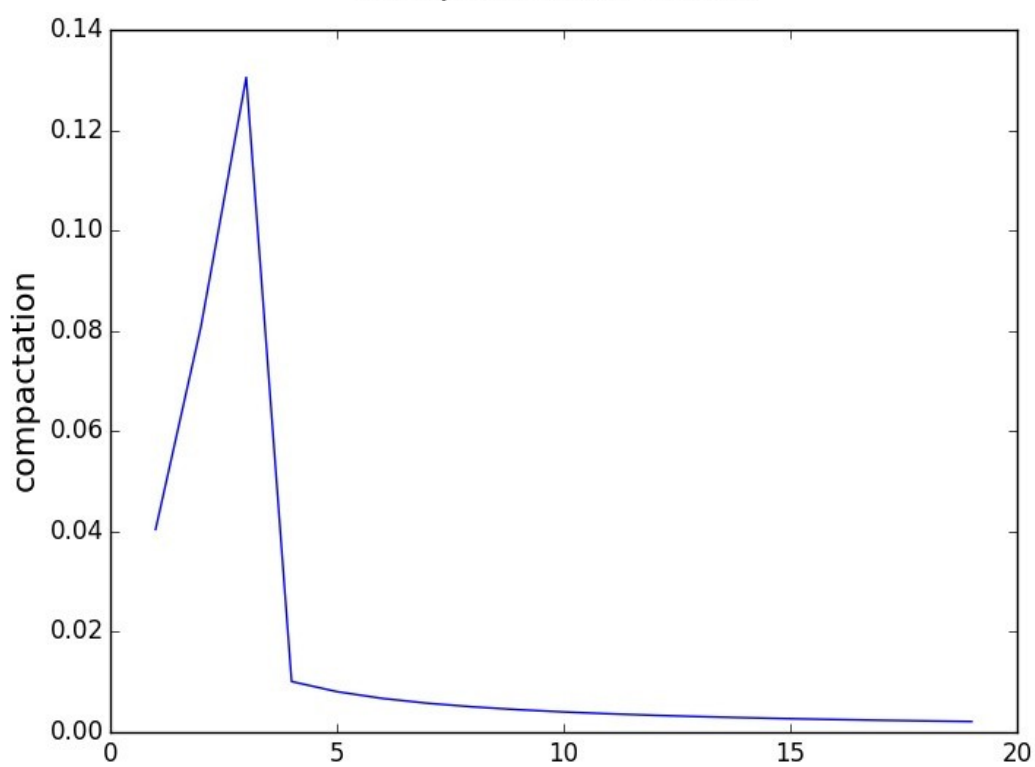
- Inicialización óptima

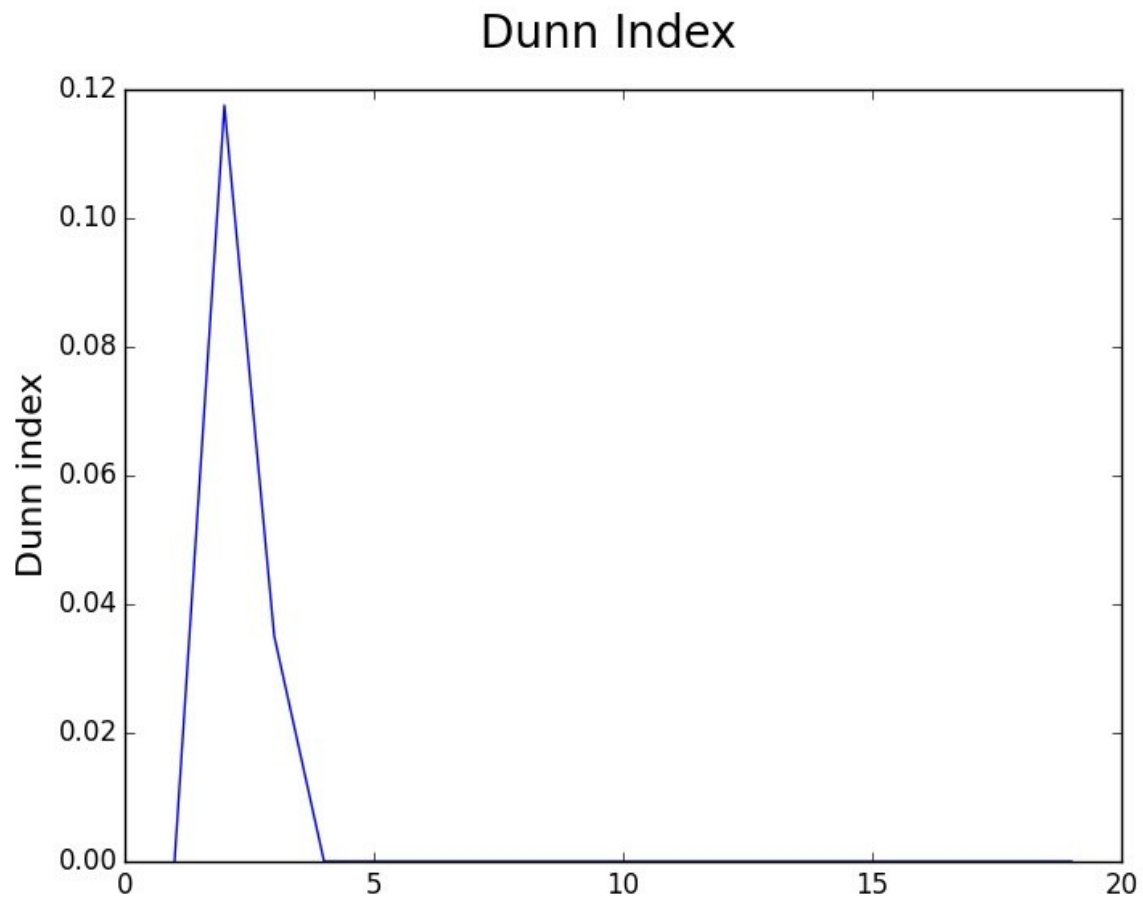
-K 1:20

Compactation Index



Compactation Index





Los resultados de la inicialización aleatoria muestran dos picos en los índices de calidad: uno para $k=5$ y otro para $k=10$ (número real de clases). El primero se repite para la inicialización óptima. Como siempre, la inicialización óptima ha convergido antes a una clusterización de calidad alta. Que ambos converjan en $k=5$ podría explicarse porque ciertos números manuscritos guardan similitud por pares como el 2 y el 5 o el 1 y el 7.

1.2 Secuencial

1.2.1 Marco teórico

El algoritmo de Clustering Secuencial tiene los siguientes parámetros críticos: un umbral a partir del cual, si la distancia de un dato a los clusters existentes supera ese umbral se crea un nuevo cluster, y un número máximo de clusters a crear.

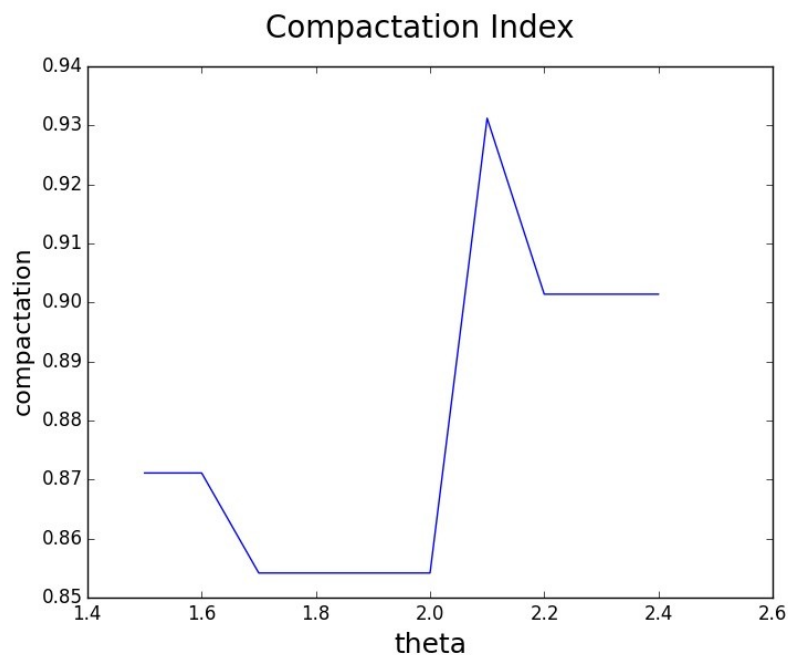
Inicialmente se comienza con un solo cluster y se van creando de la forma antes mencionada hasta que se han creado todos los clusters permitidos.

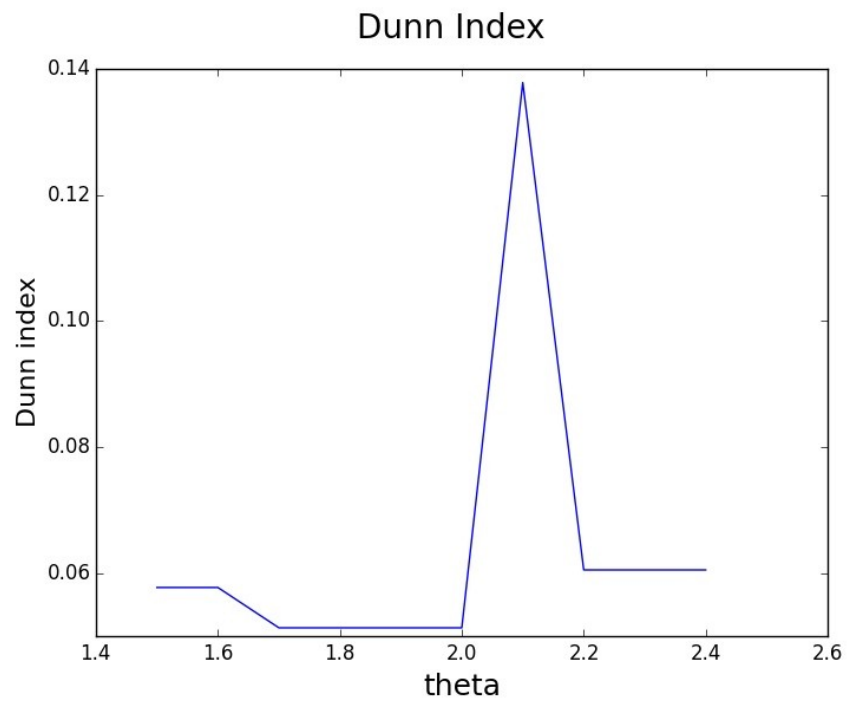
1.2.2 Resultados

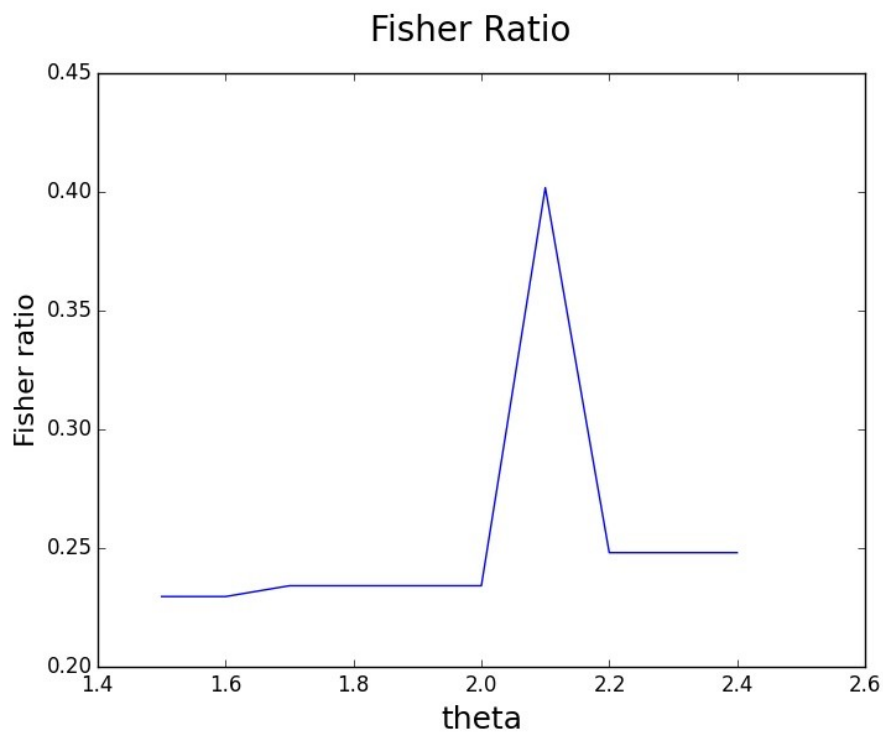
Se ha realizado el estudio del índice de compactación, el índice de Dunn y el índice de Fisher para todos los data sets de la asignatura para los distintos valores del parámetro crítico theta: el umbral a partir del cual se crea un nuevo cluster. Como número máximo de clusters se le ha pasado el número real de clases.

Se han probado los data sets con distintos rangos de theta en los que se manifestaban claras diferencias en la calidad de los índices, probando valores cercanos a la interdistancia de los puntos.

a) Iris

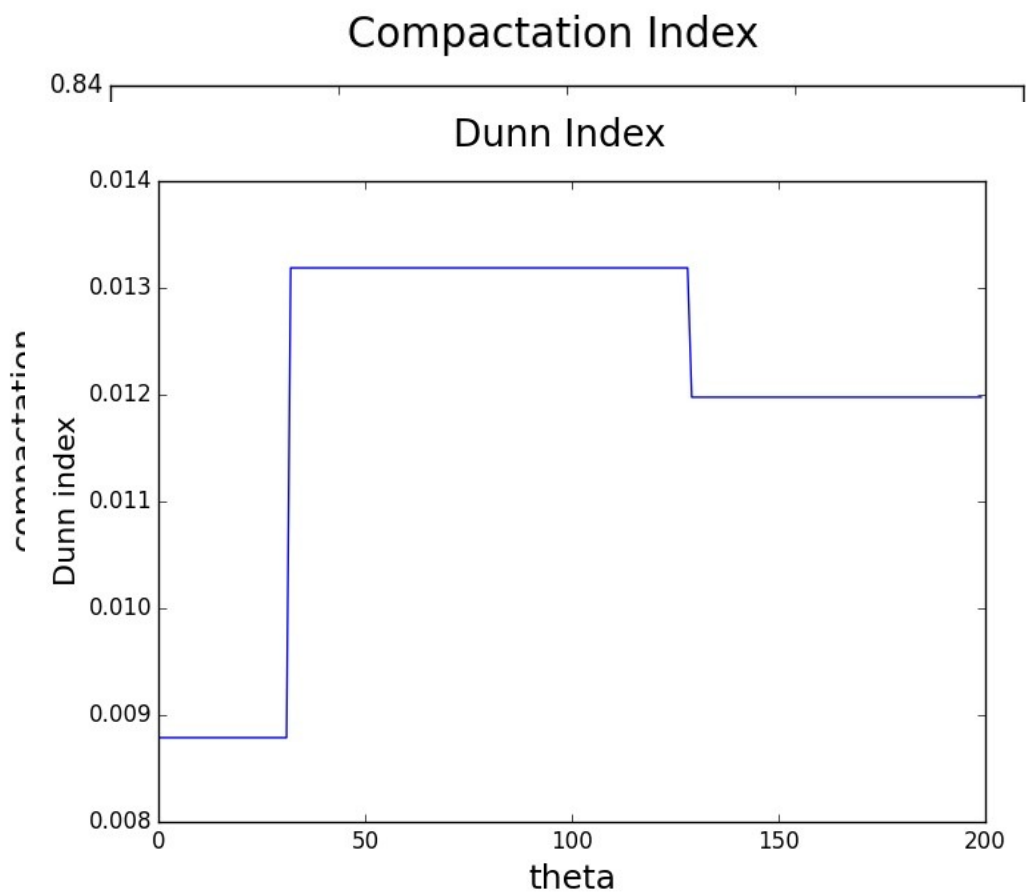


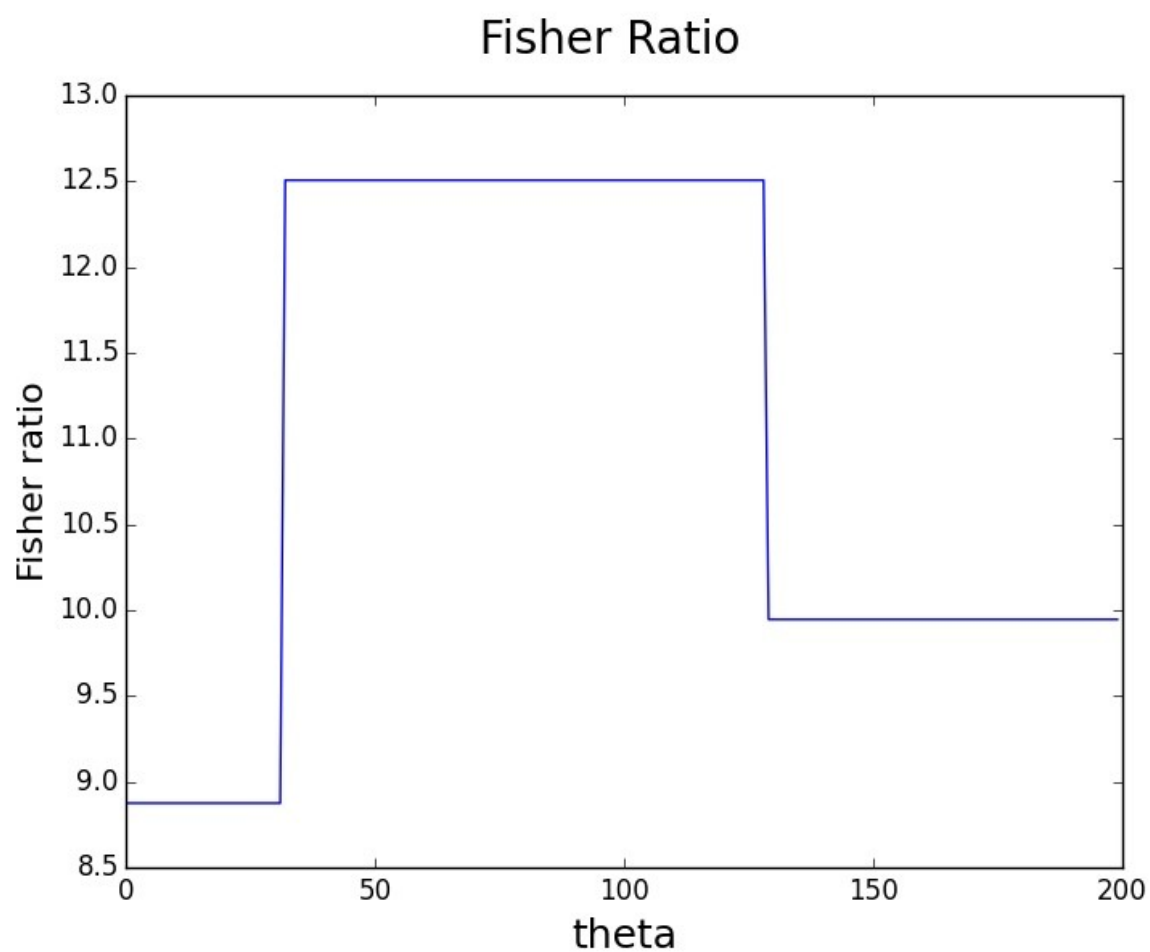




Las tres gráficas sitúan el valor óptimo de theta en 2,1 .

b) Wine

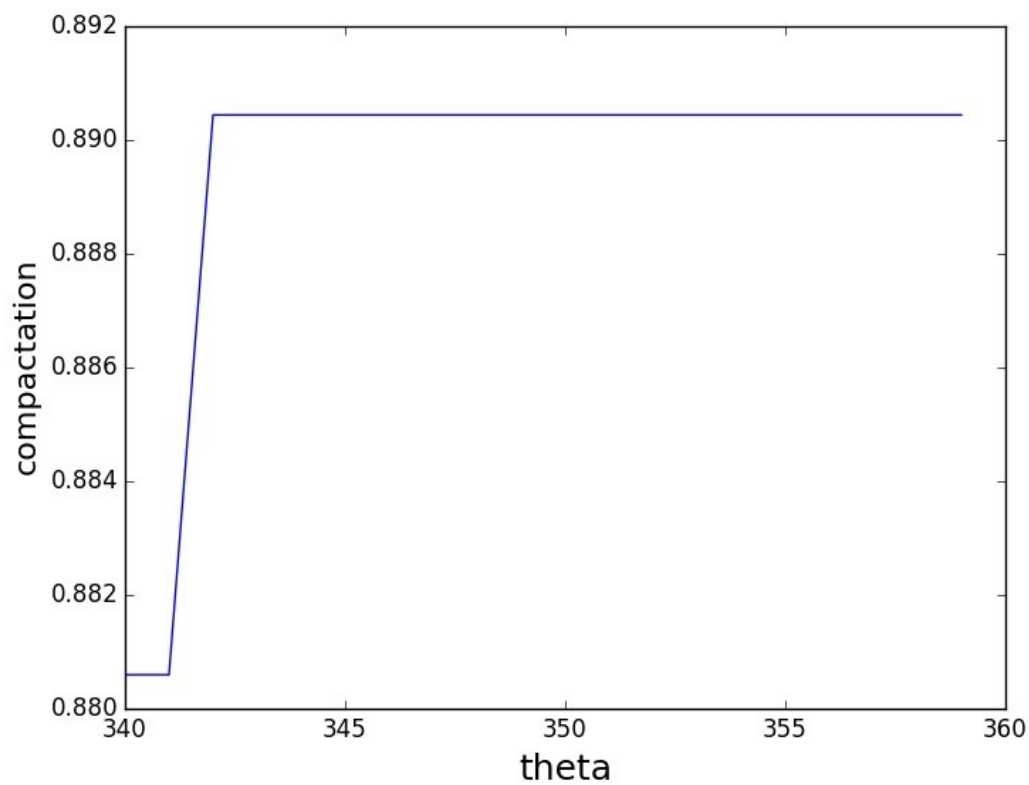




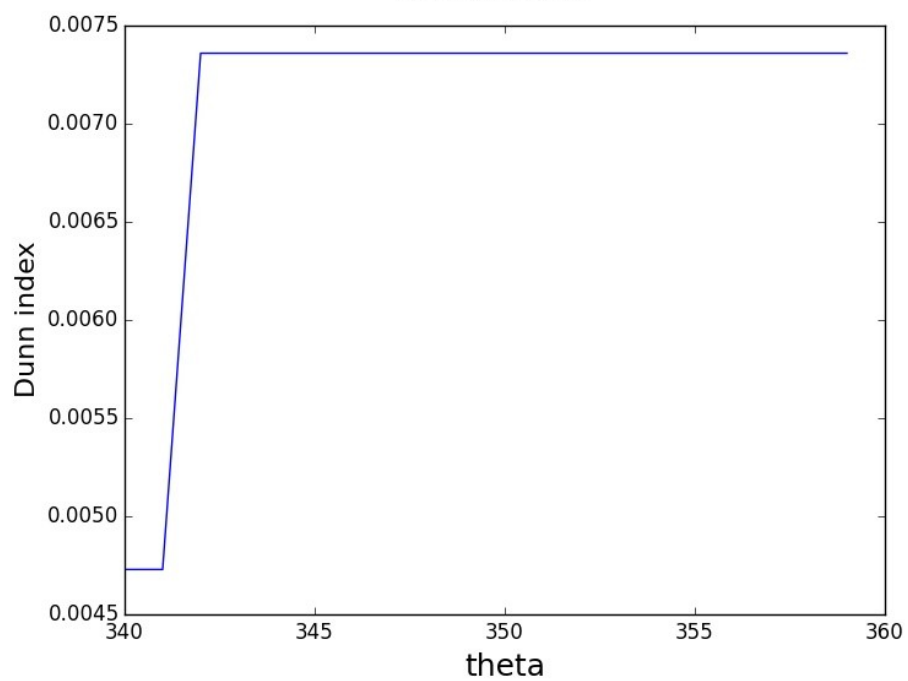
En este data set vemos la peculiaridad de que el valor óptimo de θ corresponde a todo un intervalo, situando el parámetro entre 35 y 135.

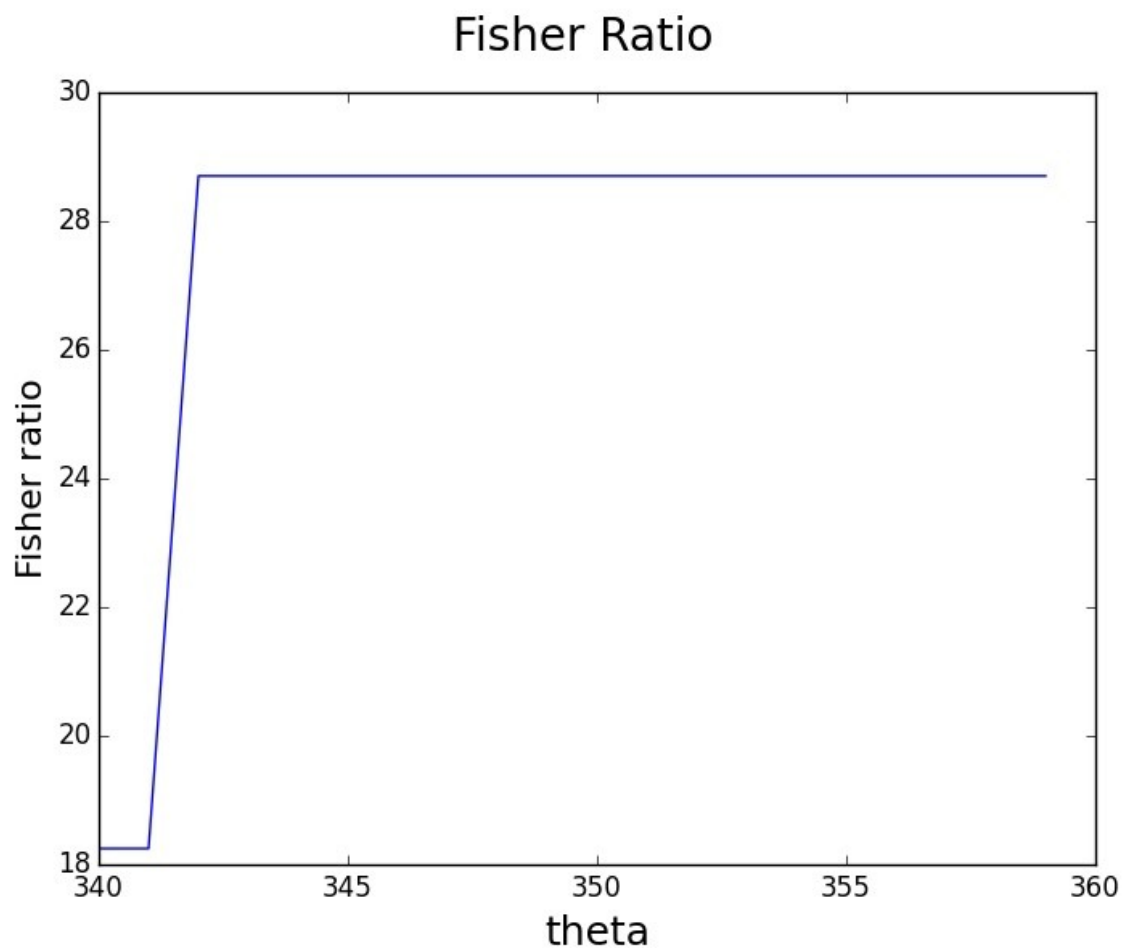
c) Cáncer

Compactation Index



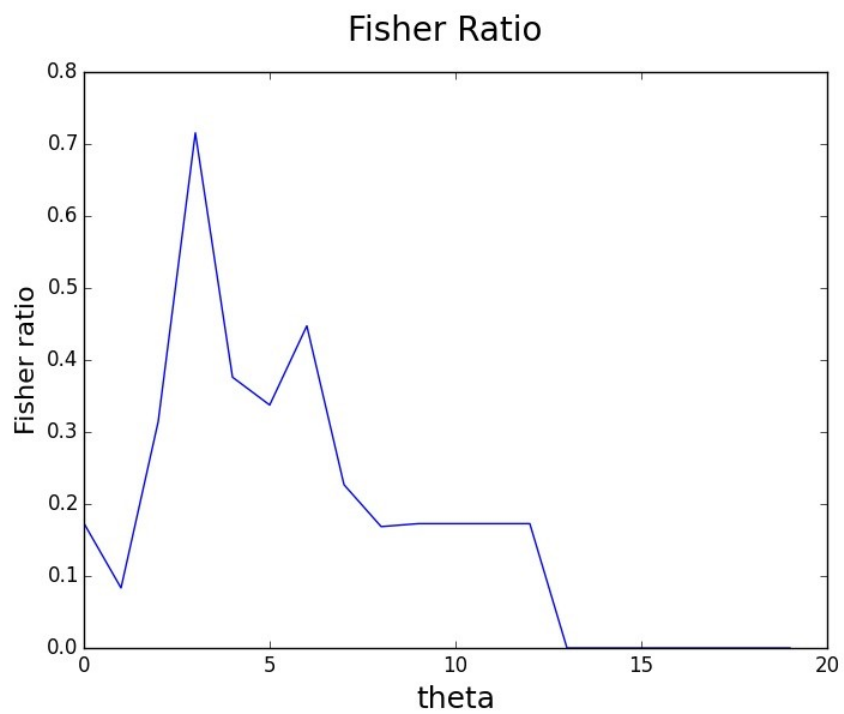
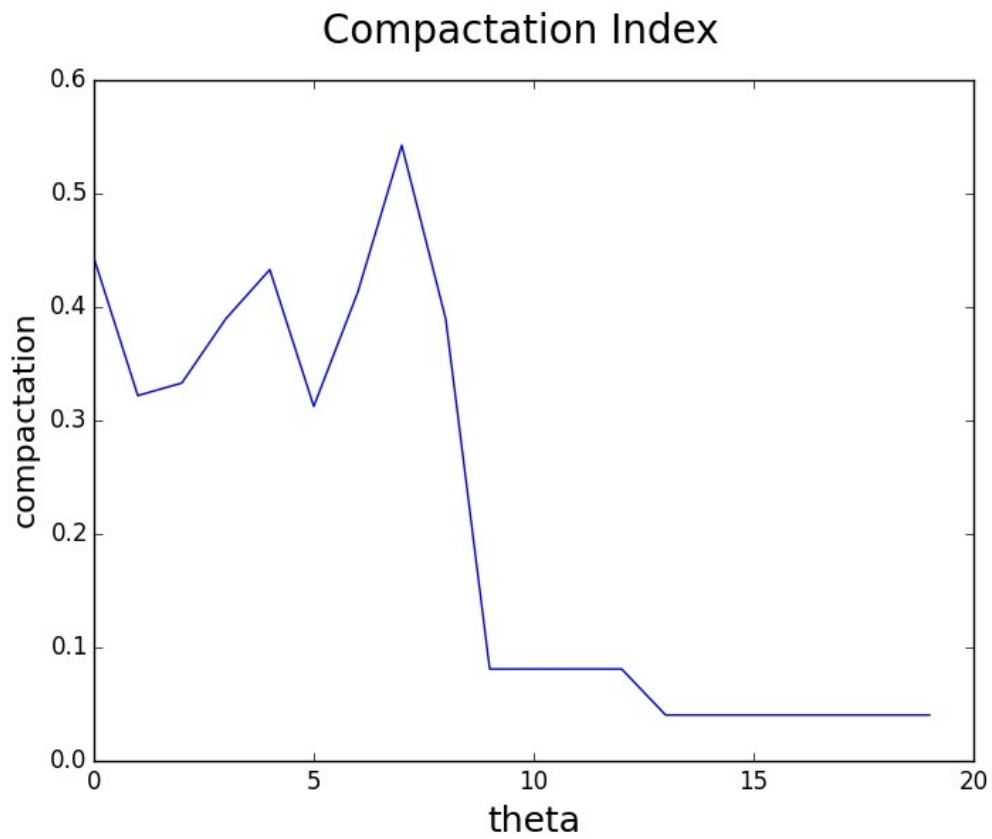
Dunn Index

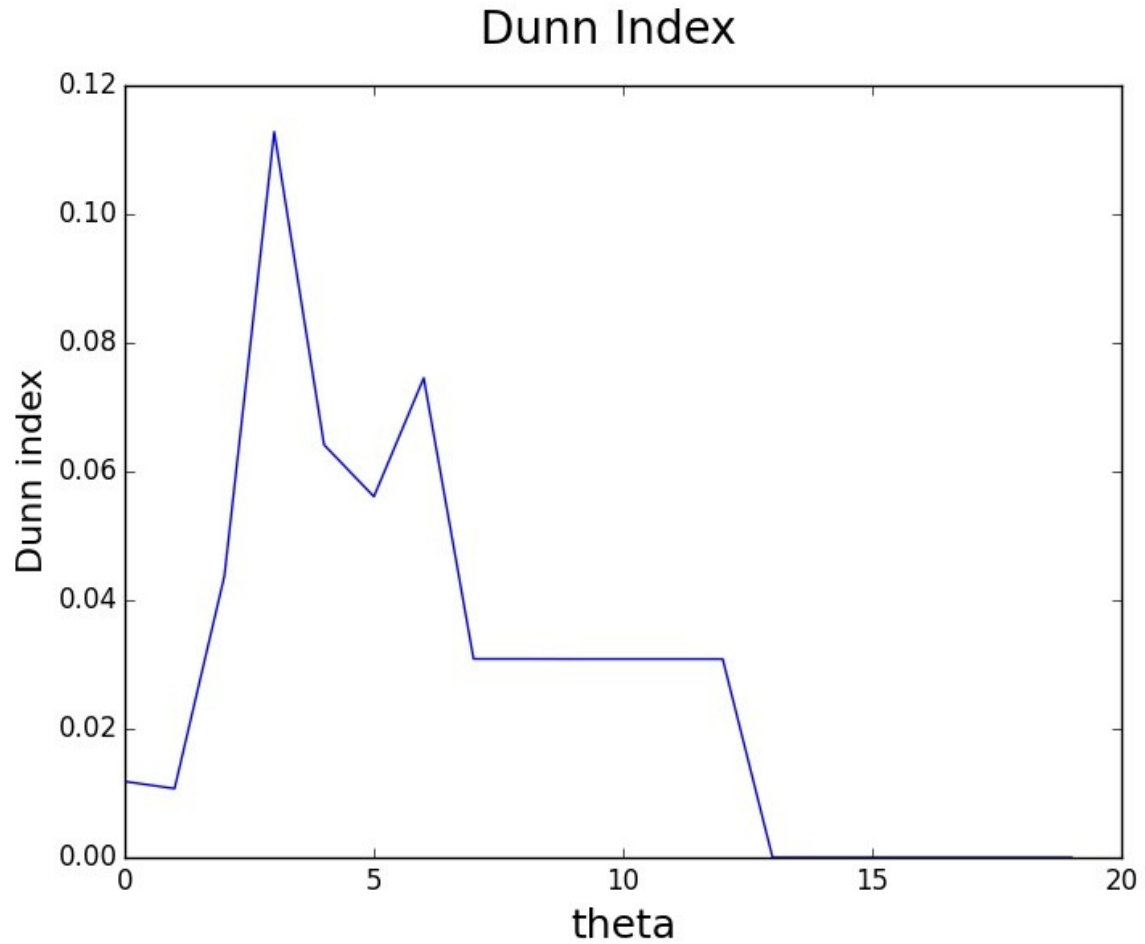




El valor a partir del cual existe un cambio claro de los resultados en función del umbral es $\theta=341$, luego los resultados permanecen estables.

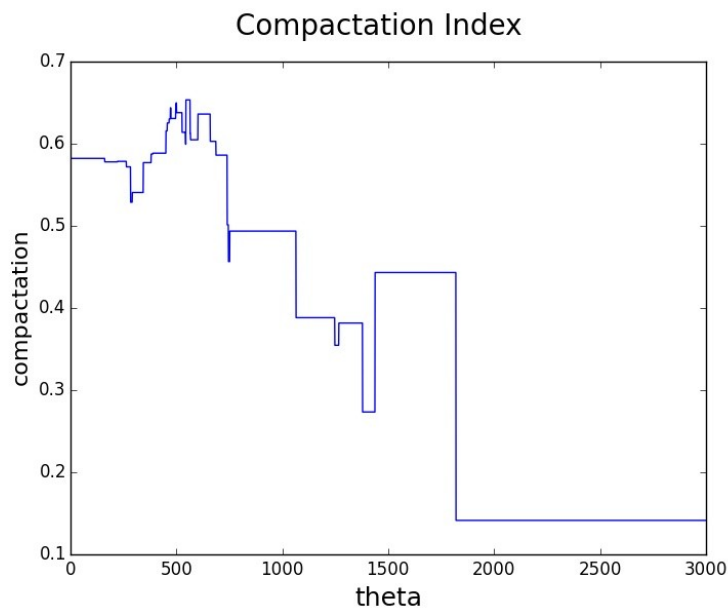
d) Isolet

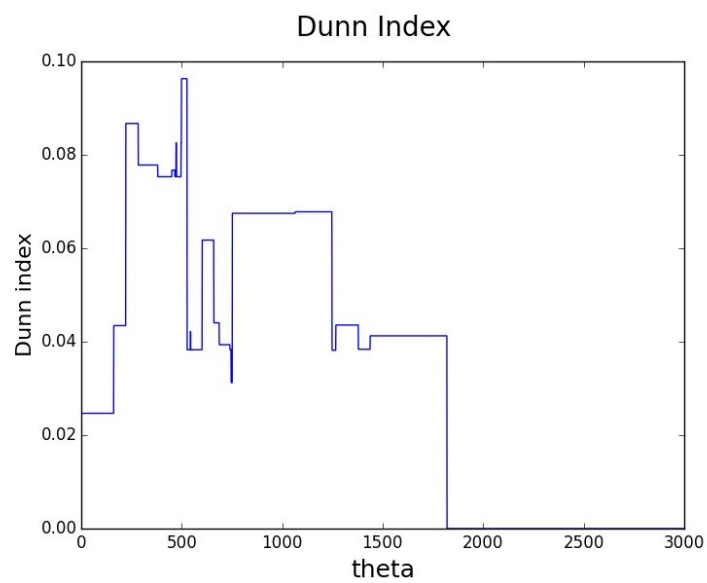
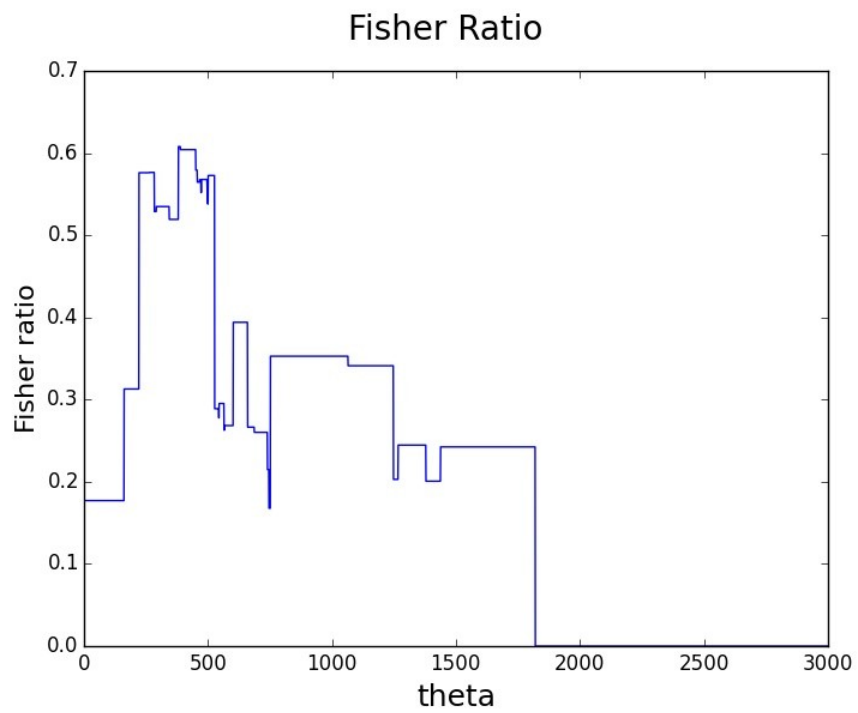




Las tres gráficas sitúan un máximo de calidad en $\theta=4$. La compactación es también alta para $\theta=7$ como resultado, seguramente, de que se crean más clusters y por lo tanto sus miembros son más parecidos entre sí.

e) MNIST





Debido a la enorme cantidad de datos de este data set y probablemente también a la variedad interna de clases tan diferentes como dígitos manuscritos los resultados para el umbral no son tan claros como en los casos anteriores. Sin embargo, las tres gráficas presentan un máximo para $\theta=500$.

2. Aprendizaje supervisado

Este tipo algoritmos trabaja con un conjunto de datos etiquetados de entrenamiento de los cuales va iterativamente aprendiendo hasta completar su aprendizaje, momento en el que se aplicarán sobre los datos reservados para test.

2.1 Perceptrón

2.1.1 Marco teórico

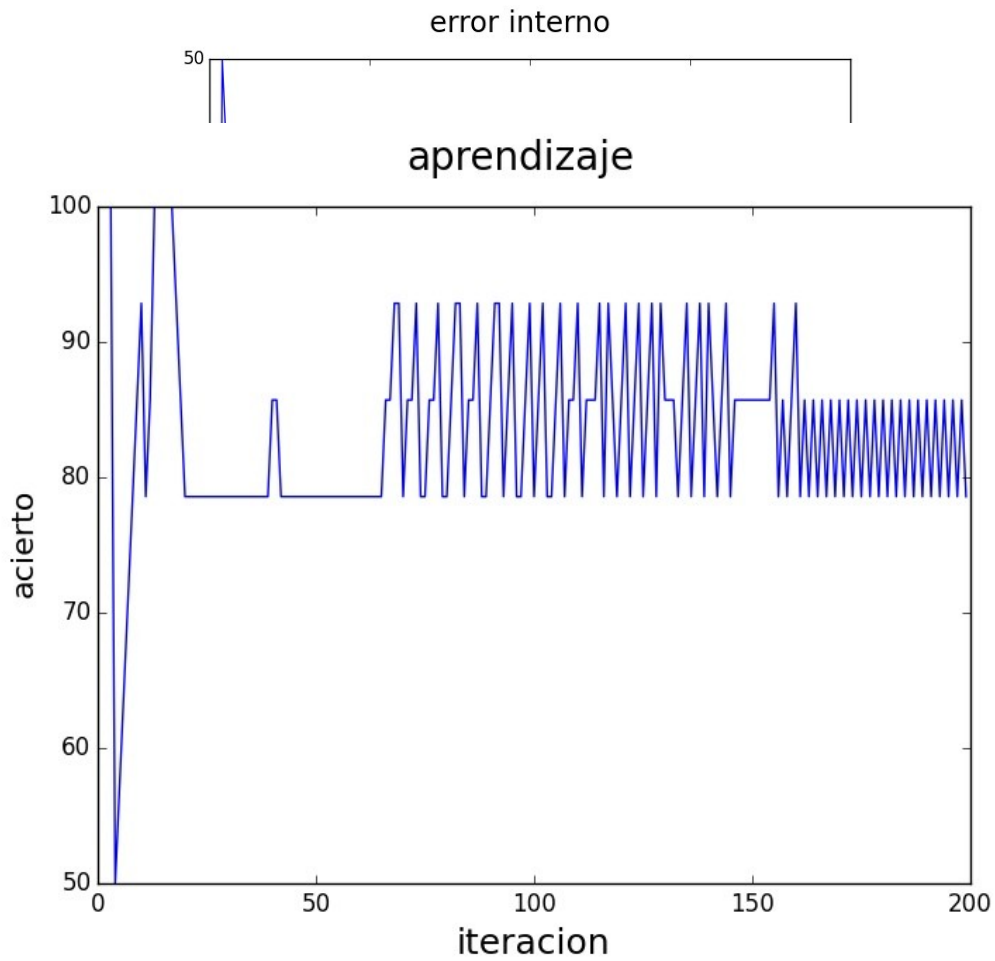
Para la evaluación y entrenamiento del algoritmo dividiremos primero el data set en un 90% de datos para entrenamiento y el 10 % de datos para test (realizaremos un k-fold con $k=10$). Dentro de ese 90% utilizaremos el 90% para aprendizaje efectivo y el 10% para lo que llamaremos validación interna.

Iniciamos las funciones discriminantes en la denominada posición de indiferencia, en nuestro caso con todos los pesos a cero. A continuación clasificaremos los datos etiquetados del conjunto de aprendizaje efectivo utilizando estas funciones discriminantes. En caso de que la clasificación falle, se incrementará en un factor μ ($1/k$) los pesos correspondientes a la etiqueta correcta y se disminuirán en la misma proporción los no correspondientes. Este proceso se repetirá hasta que la validación interna converja a un valor pequeño y estable. Una vez finalizado, ya tenemos el clasificador entrenado para poder aplicarlo a los datos de test.

En el informe de la práctica se incluyen las curvas de validación interna y aprendizaje para justificar el número de iteraciones del proceso escogido.

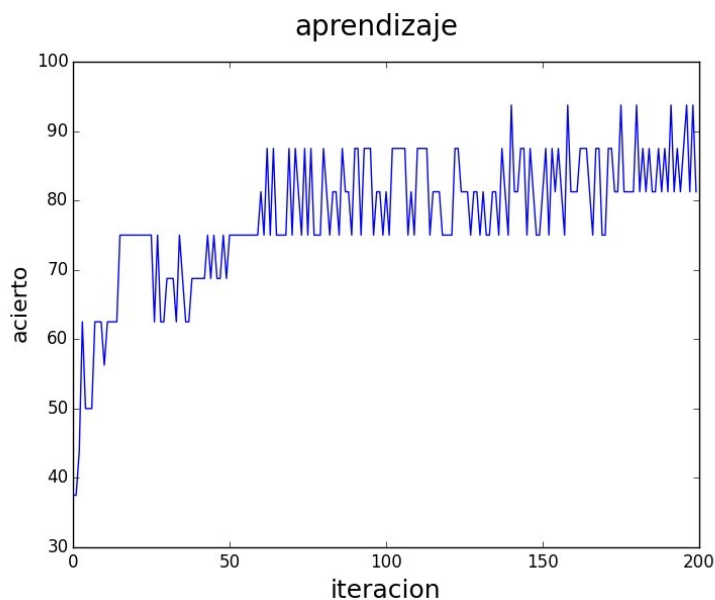
2.1.1 Resultados

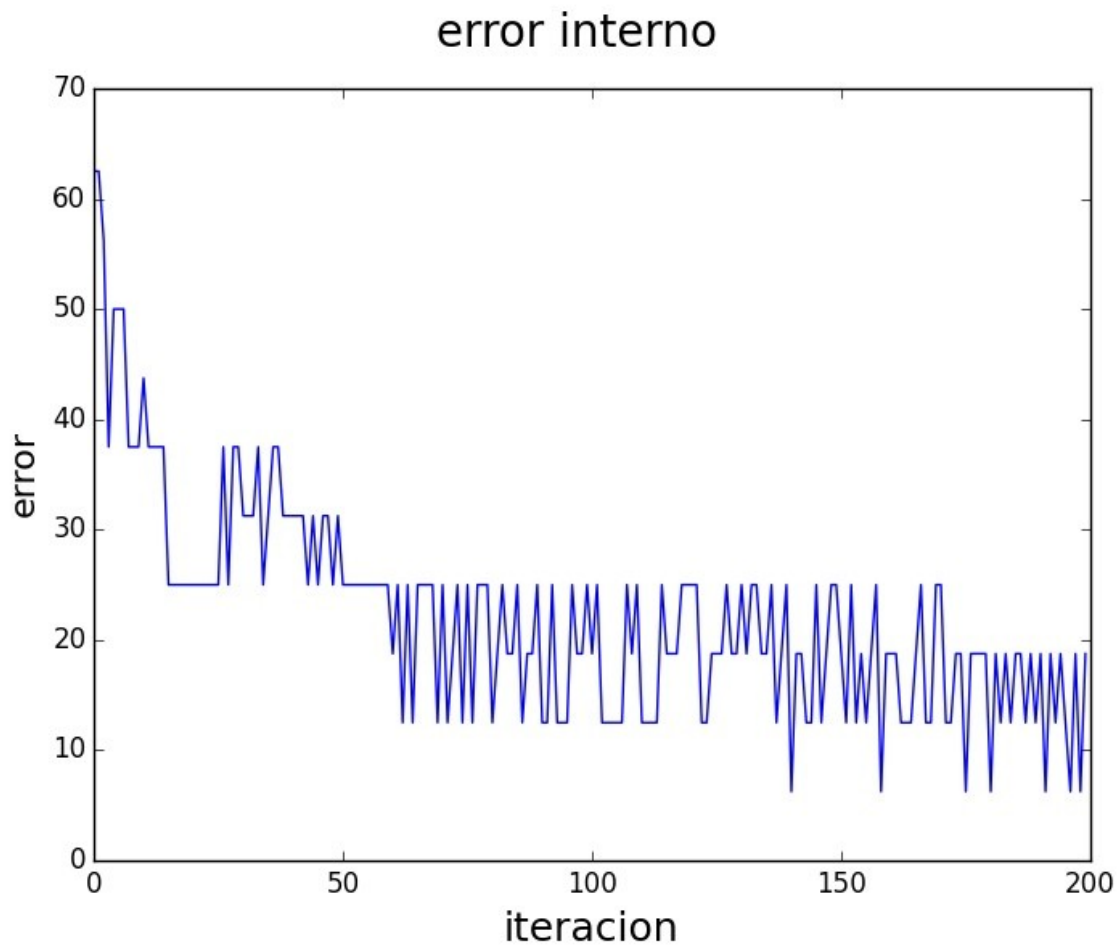
a) Iris



Vemos que el aprendizaje se estabiliza bastante (no totalmente porque los conjuntos de validación interna son muy pequeños y un solo fallo provoca un gran error) a partir de la iteración 20. El porcentaje de error final ha sido del 7,1 %.

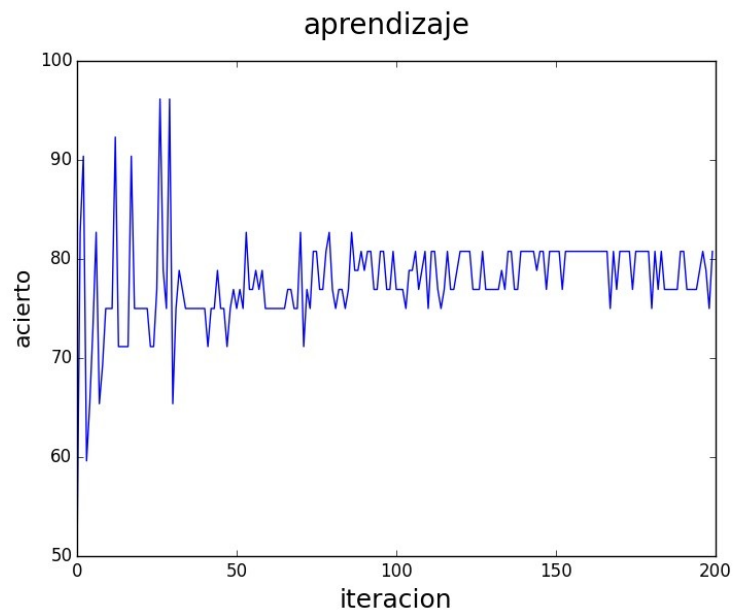
b) Wine

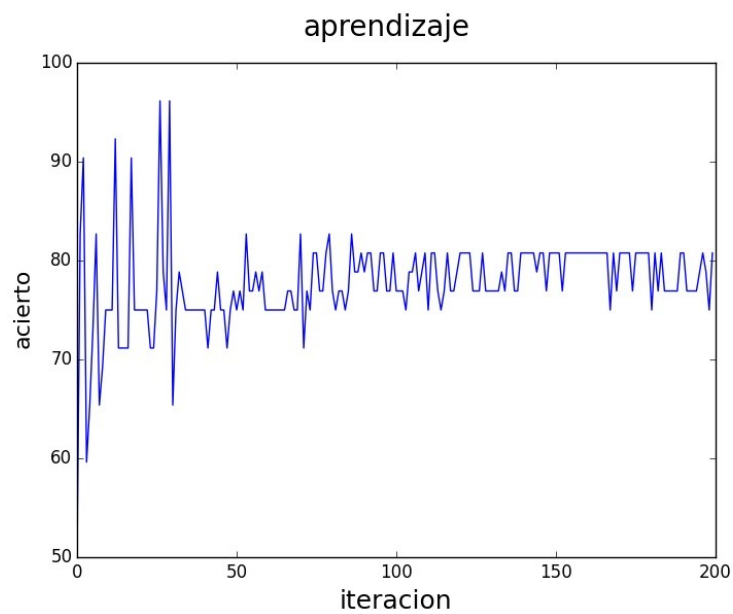




Aprendizaje y error se estabilizan en 50 iteraciones, valor para el cual se ha logrado una tasa de error en el test del 2,7%.

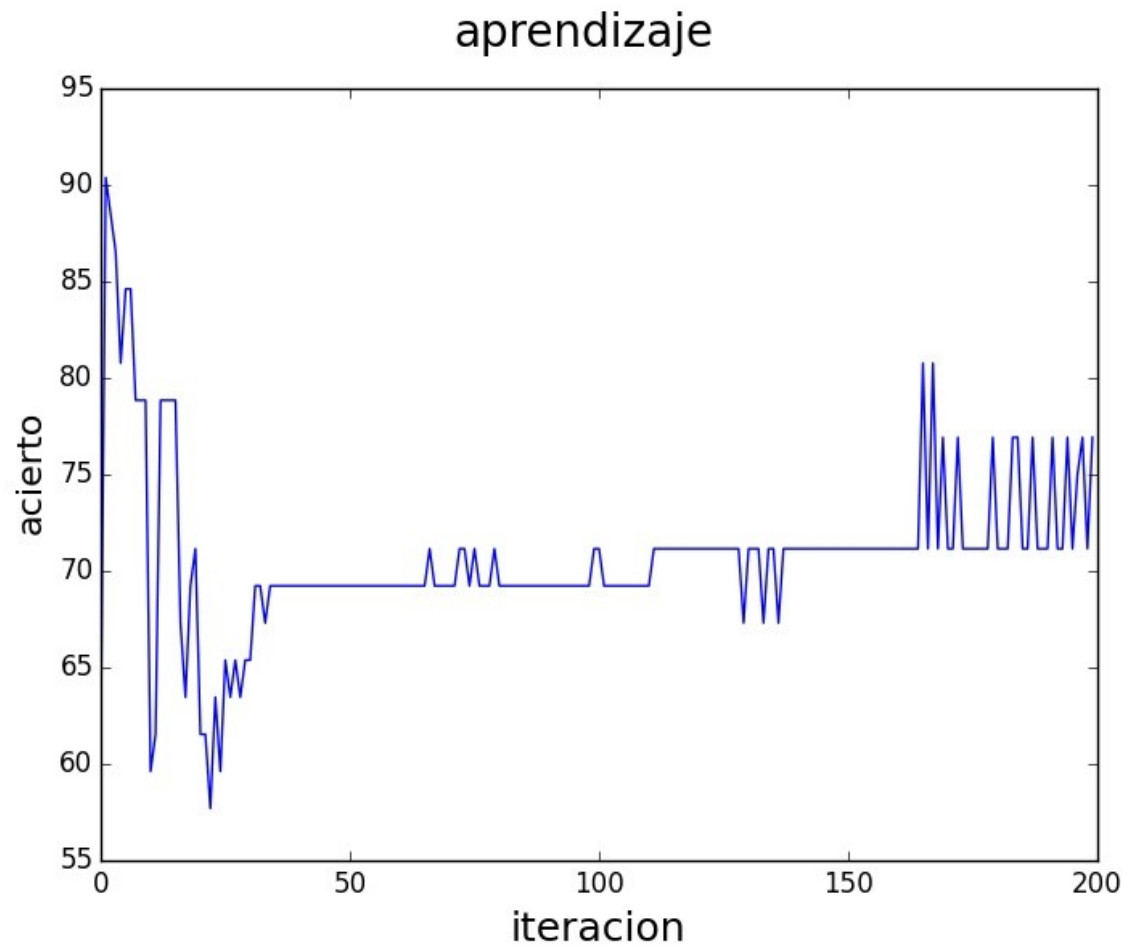
c)Cancer

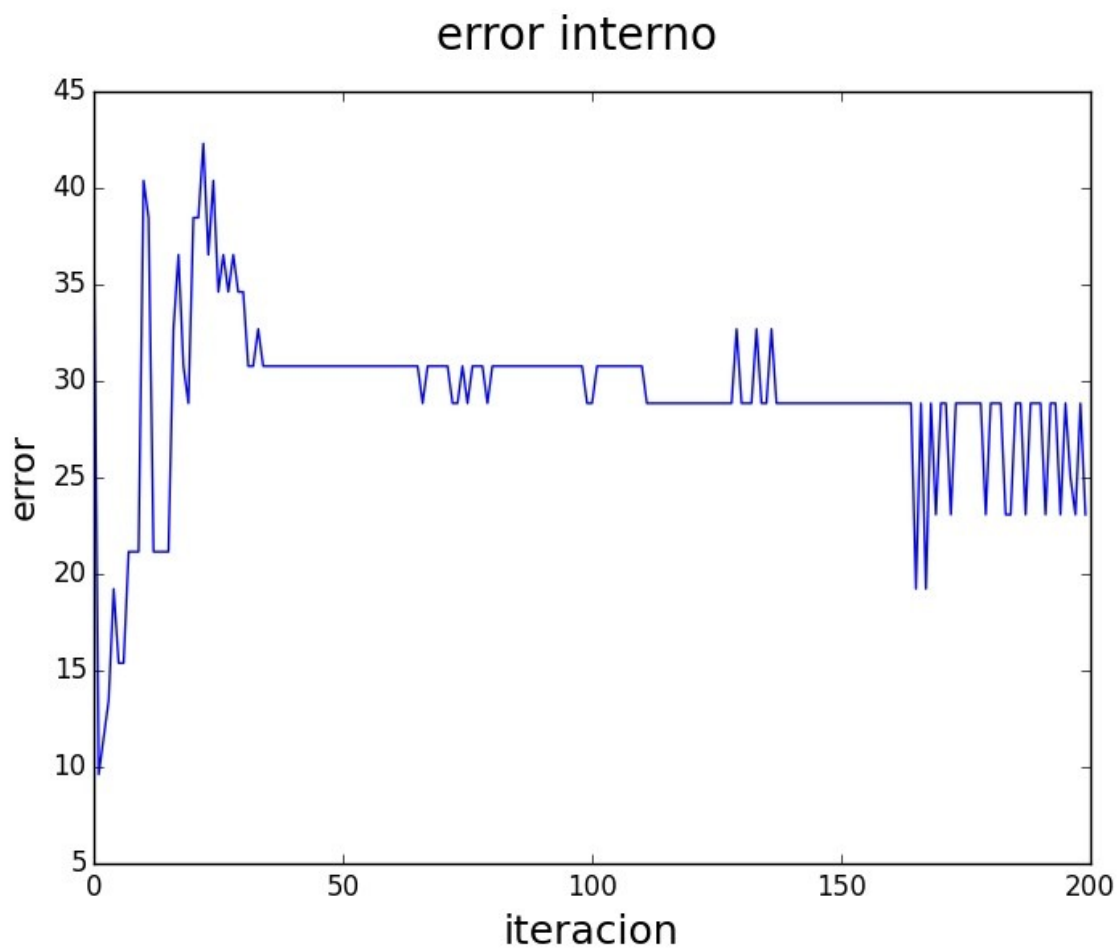




La estabilización del número de errores y aciertos llega sobre la iteración 80 con un porcentaje de error del 2,28%.

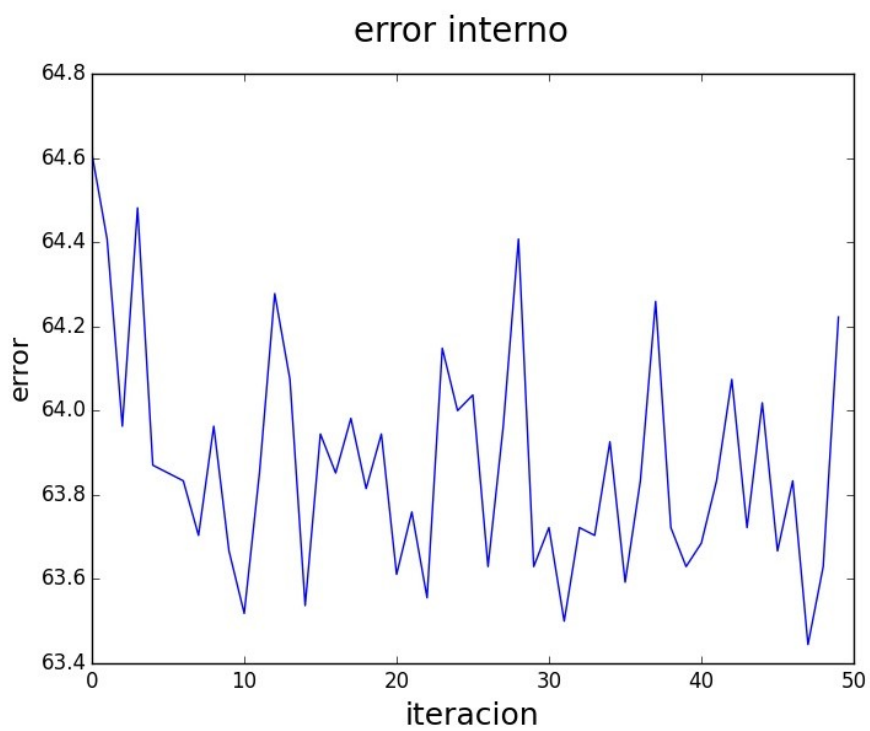
d) Isolet

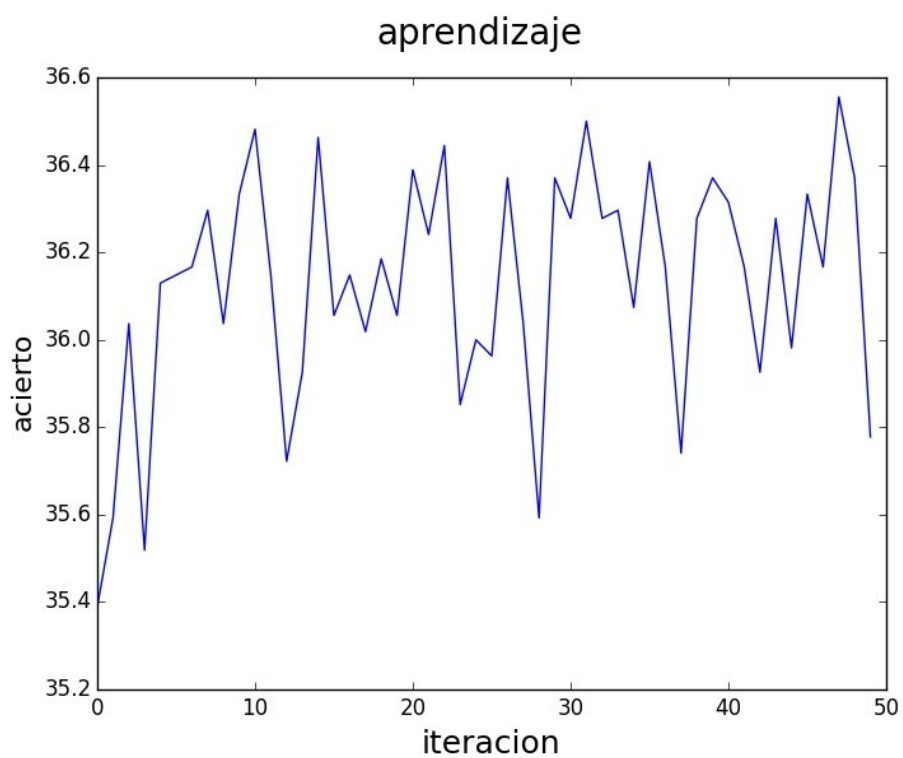




Se estabiliza en la iteración 45 con un error del 5,1%.

e)MNIST





No se logra gran estabilidad aunque es cierto que los porcentajes de variación tampoco son excesivamente grandes. Un número mayor de iteraciones era impracticable en cuanto a tiempos de ejecución.

2.2 Ensemble

2.2.1 Marco teórico

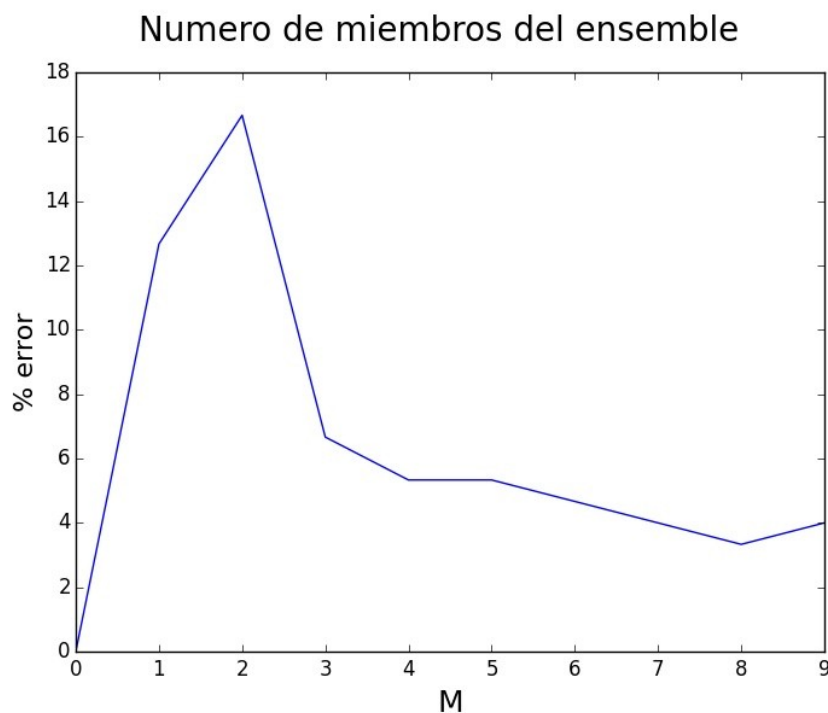
Un ensemble podría definirse sintéticamente como un conjunto de clasificadores actuando conjuntamente para decidir la etiqueta de un dato, normalmente, mediante la regla de la mayoría de votos. Distinguimos en esta práctica dos técnicas para crear ensambles: bagging, cuando el conjunto de entrenamiento para cada miembro del ensemble se obtiene de forma independiente mediante una elección al azar con reemplazamiento dentro del conjunto etiquetado, y boosting, en el que cada miembro del ensemble se obtiene testeando en el data set completo y entrenando en el conjunto de datos mal etiquetados por el miembro del ensemble anterior.

2.2.2 Resultados

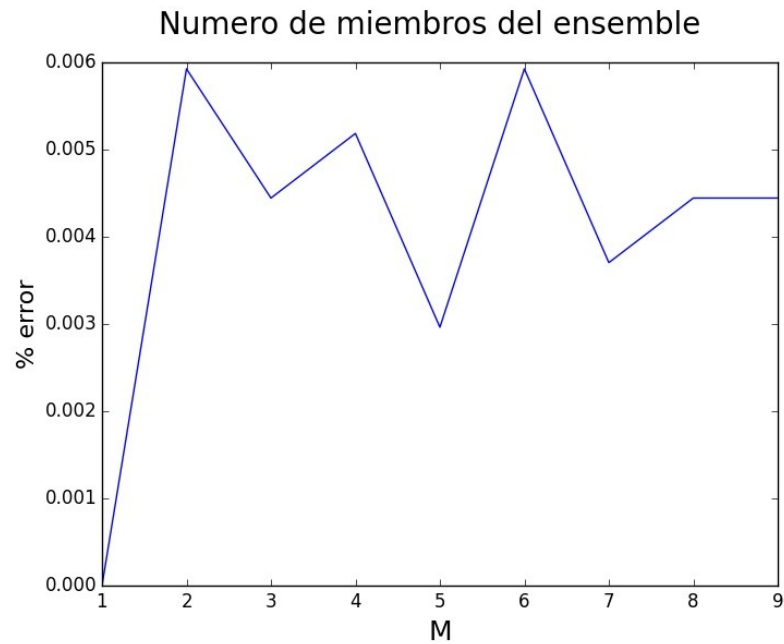
El parámetro crítico a optimizar en la práctica será el número de miembros del ensemble M .

a) Iris

-Bagging



-Boosting

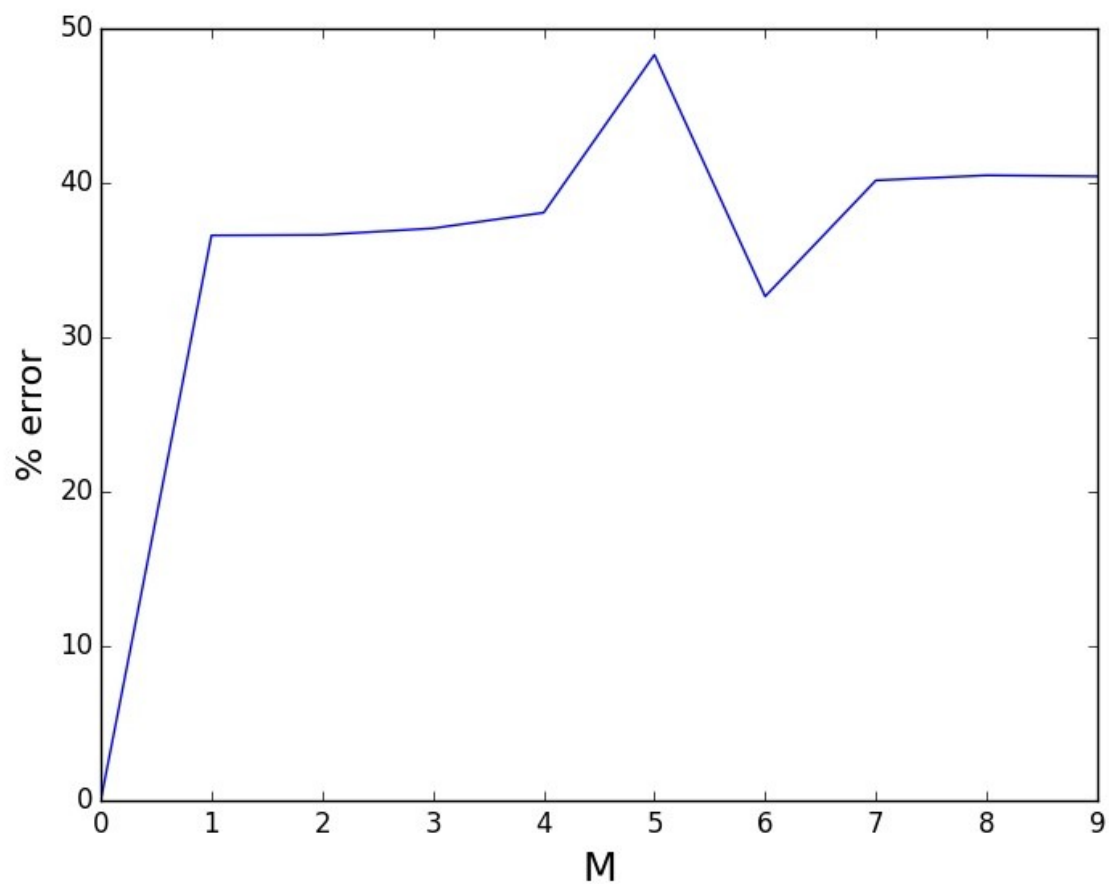


Mediante boosting el mínimo de error se alcanza claramente para ensambles de 5 miembros mientras que según la técnica de bagging, aunque 5 también da uno de los errores más bajos, con 8 miembros el error sería todavía menor.

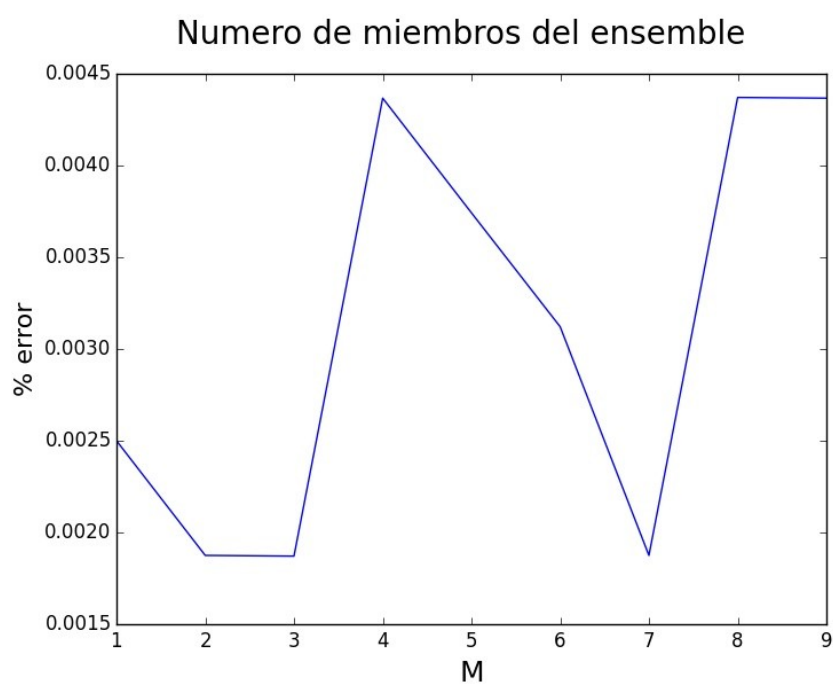
b) Wine

-Bagging

Numero de miembros del ensemble



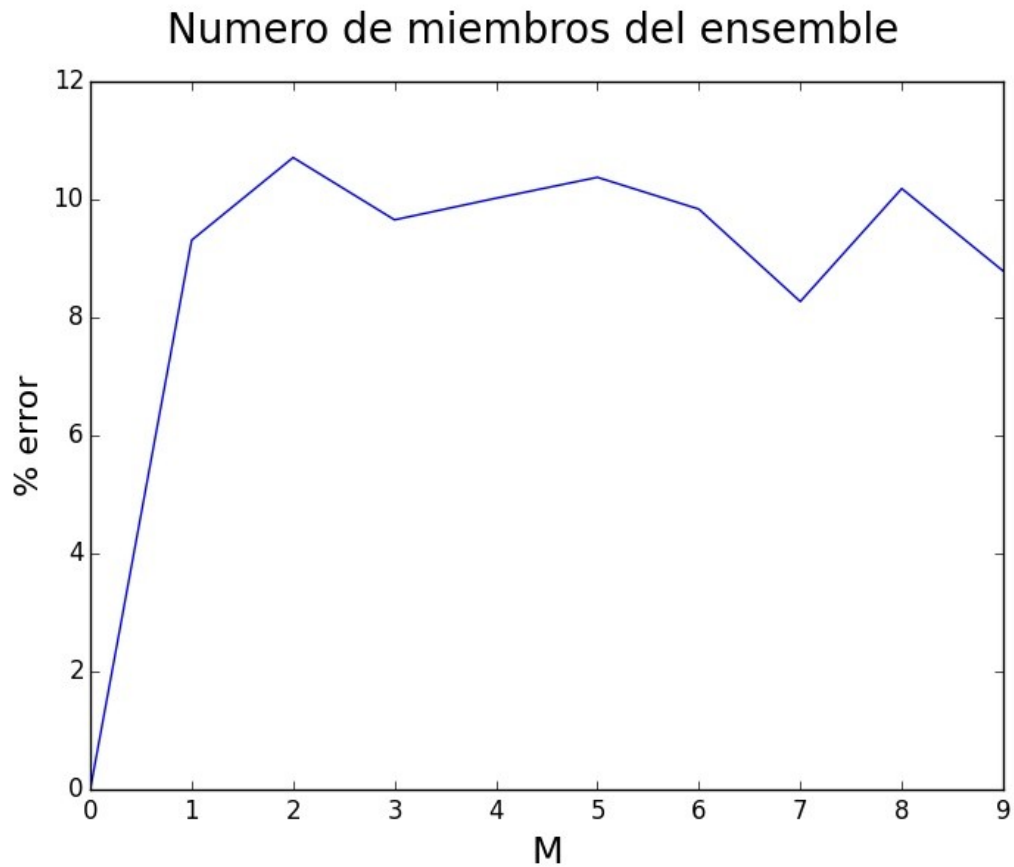
-Boosting



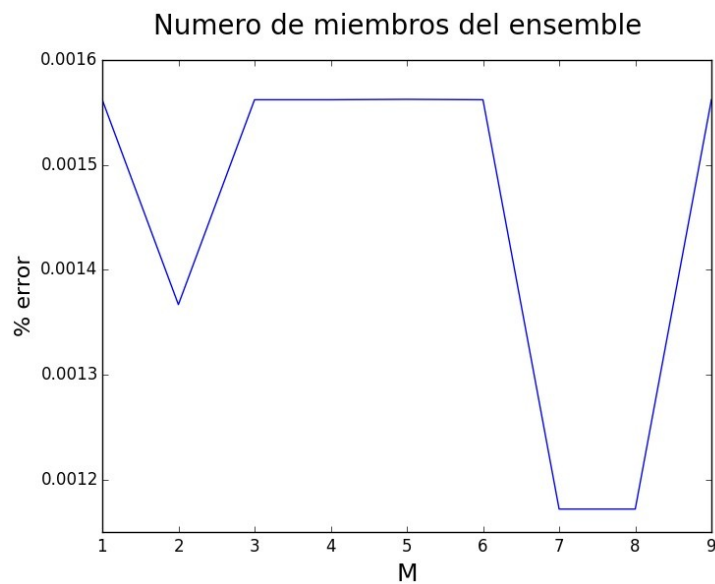
Según la técnica de bagging, para este data set la solución más óptima seían ensambles de 6 miembros y según boosting de 7.

c)Cancer

-Bagging



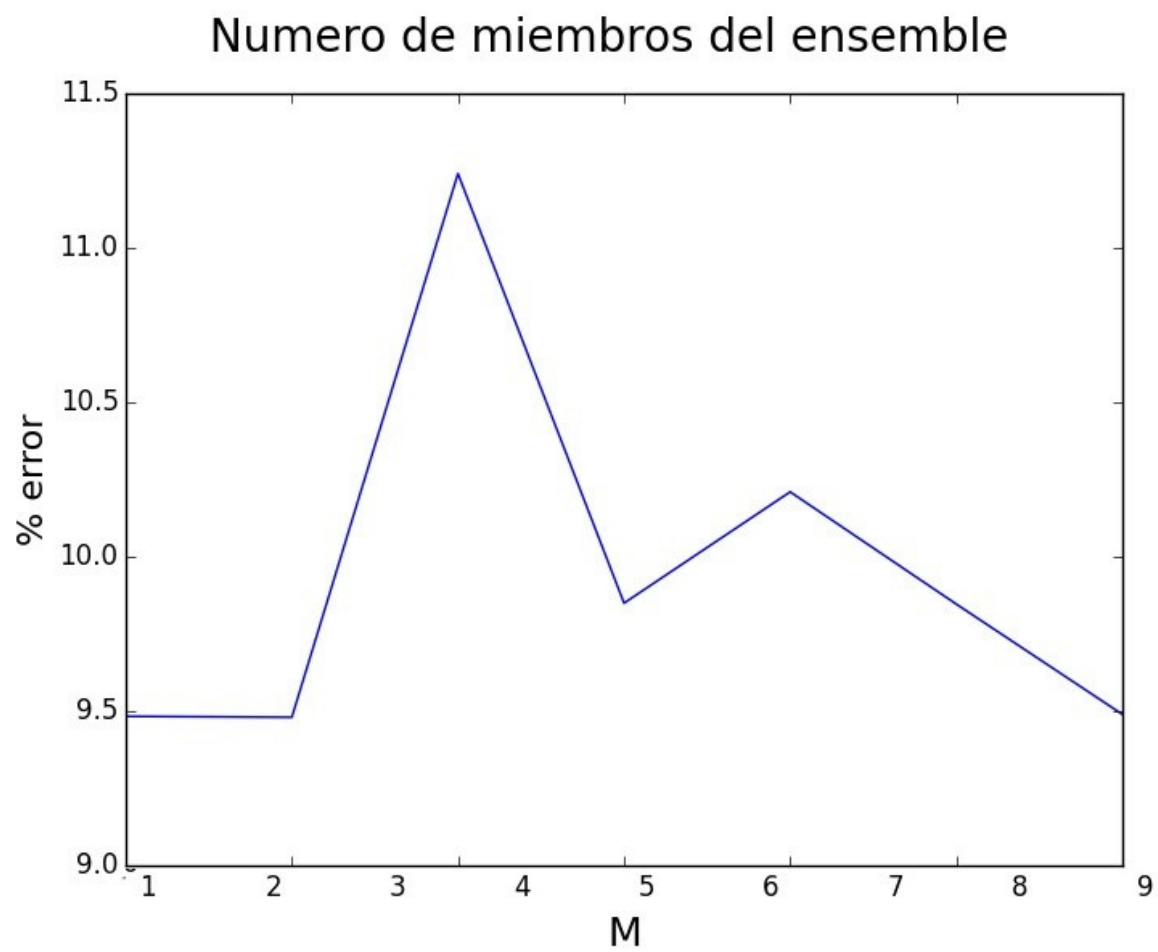
-Boosting



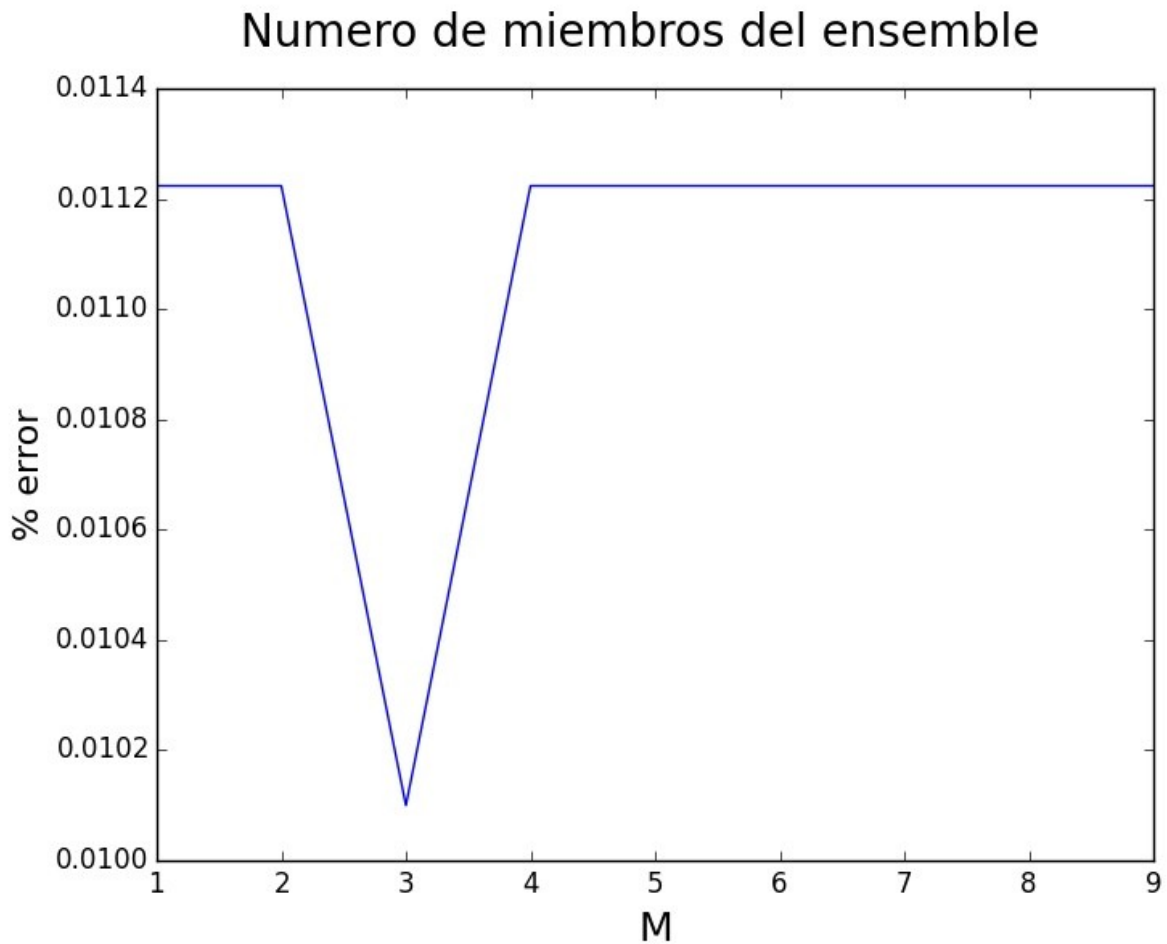
En este caso ambas técnicas coinciden en situar el mínimo error en $M=7$.

d) Isolet

-Bagging



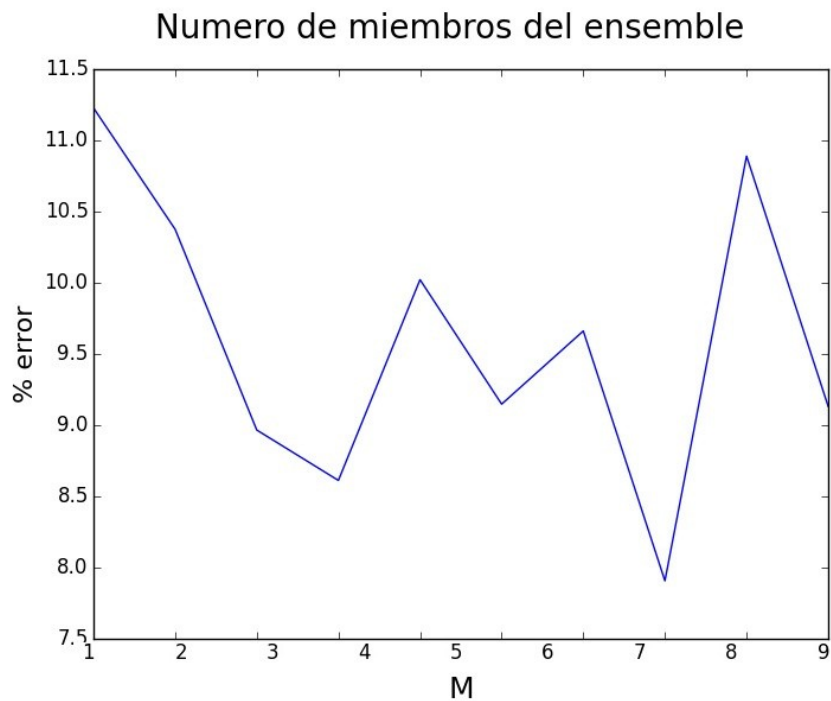
-Boosting

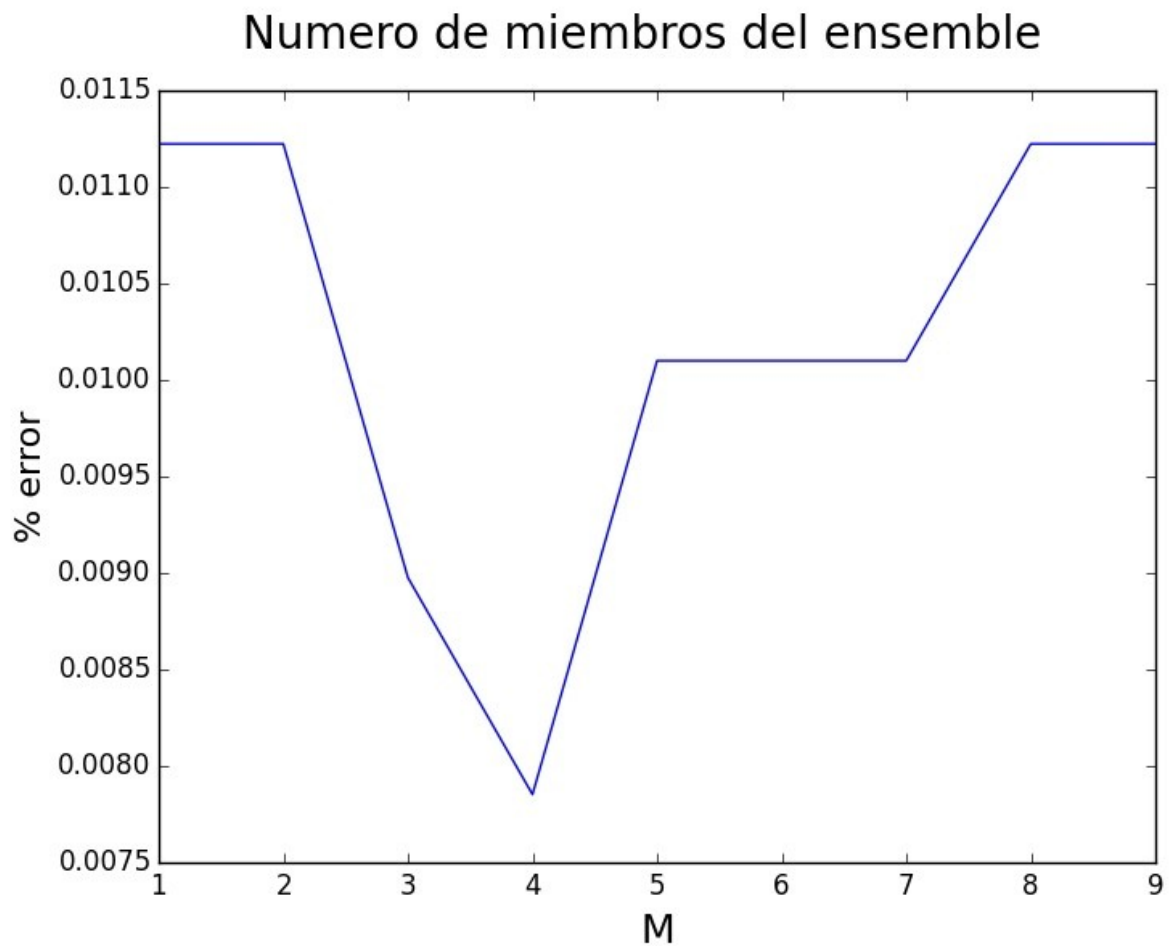


En este caso es mediante la técnica de boosting que se obtiene una respuesta más clara optimizando M en el valor 3.

e) MNIST

-Bagging



-Boosting

Ambas técnicas sitúan un mínimo de error para $M=4$ que para boosting es absoluto. Bagging presenta su mínimo absoluto para $M=7$.

2.3 Reglas Granuladas

2.3.1 Marco teórico

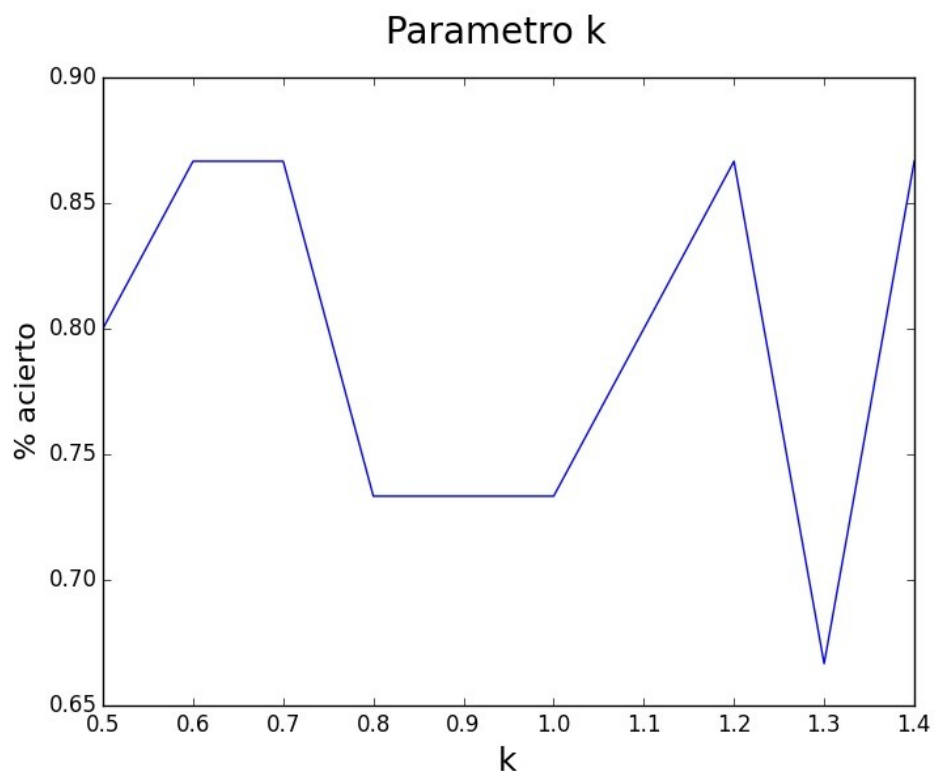
El aprendizaje en reglas graduadas consiste en clasificar los miembros del data set utilizando una serie de reglas predefinidas y jerarquizadas mediante pruebas. El proceso de creación de estas reglas consiste en primero crear todas las que sean posibles: los valores de cada parámetro se dividen en tres segmentos (bajo, medio y alto) según unos umbrales definidos por la media del parámetro, su desviación y un parámetro crítico k cercano a 1, y a continuación se forman todas las posibles combinaciones de estos segmentos para todos los parámetros y todas las etiquetas.

Se mide cuales de esas reglas son las más efectivas empleandolas en todo el data set y finalmente se escoge un subconjunto de estas para las labores de test.

2.3.2 Resultados

Se aplicará el estudio basado en reglas granuladas a todos los data sets de la asignatura analizando cuales son las reglas resultantes escogidas. También se buscará optimizar el parámetro crítico k . La creación de normas requiere de variaciones combinatorias lo que supone mucha carga a nivel computacional si el número de parámetros de los datos es demasiado alto. El número máximo de parámetros factible en tiempos de ejecución en mi equipo ha sido de 6 por lo que se han realizado técnicas de reducción de la dimensionalidad (PCA) a dicha dimensión.

a) Iris



Máximo de acierto con $k=0,6$ y $0,7$.

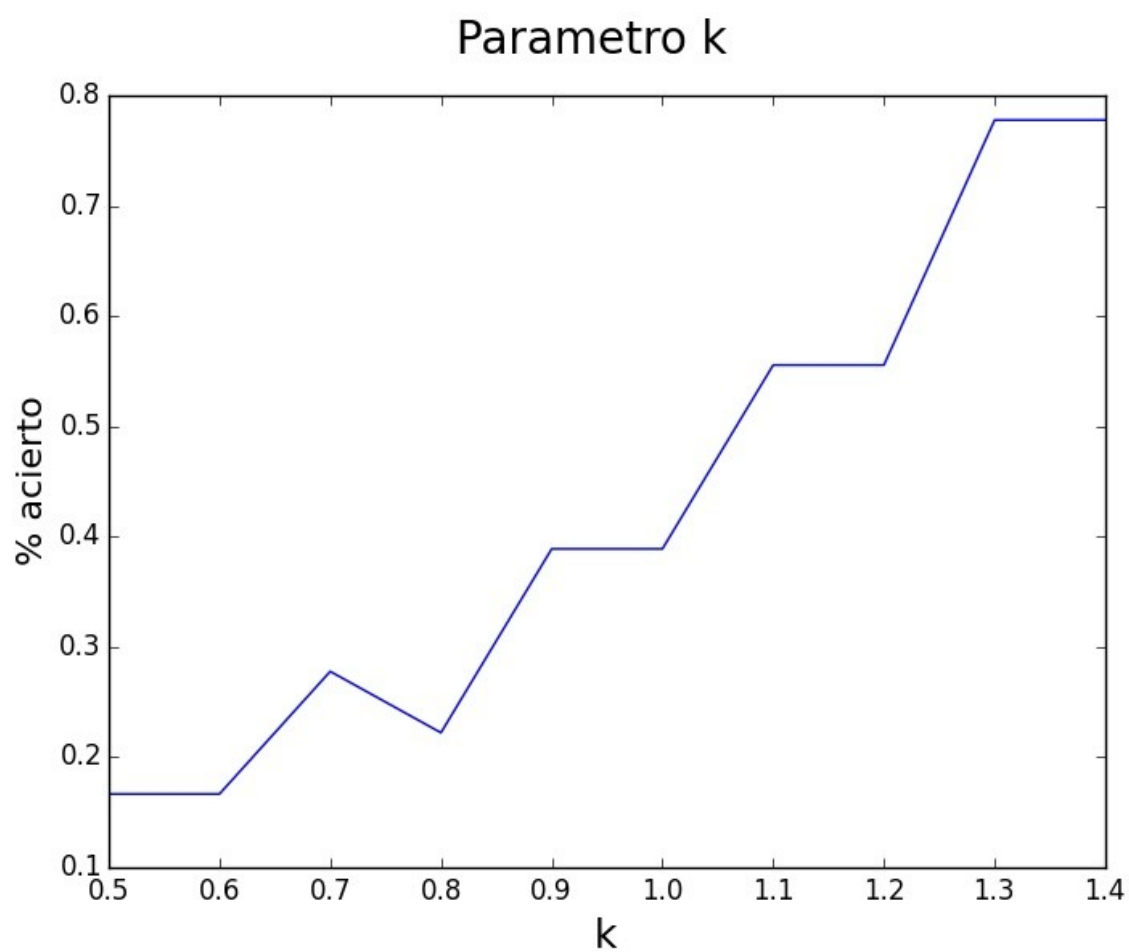
	Umbral inferior*	Umbral Superior**
Parámetro 1	5.2580057	6.44421652
Parámetro 2	2.74233015	3.365818
Parámetro 3	2.5136535	4.9989391
Parámetro 4	0.6670593	1.72553329

*Por debajo de él será considerado bajo **Por encima de él será considerado alto, medio en otro caso

	Parámetro 1	Parámetro 2	Parámetro 3	Parámetro 4	Etiqueta
Regla 1	Bajo	Alto	Bajo	Bajo	Setosa
Regla 2	Alto	Medio	Alto	Alto	Virgínica
Regla 3	Bajo	Medio	Bajo	Bajo	Setosa
Regla 4	Medio	Bajo	Medio	Medio	Versicolor
Regla 5	Medio	Medio	Medio	Medio	Versicolor
Regla 6	Medio	Alto	Bajo	Bajo	Setosa

Regla 7	Medio	Medio	Alto	Alto	Virgínica
Regla 8	Alto	Medio	Medio	Medio	Versicolor
Regla 9	Bajo	Bajo	Medio	Medio	Versicolor
Regla 10	Medio	Bajo	Alto	Alto	Virgínica

b) Wine



Porcentaje de acierto máximo para $k=1,3$.

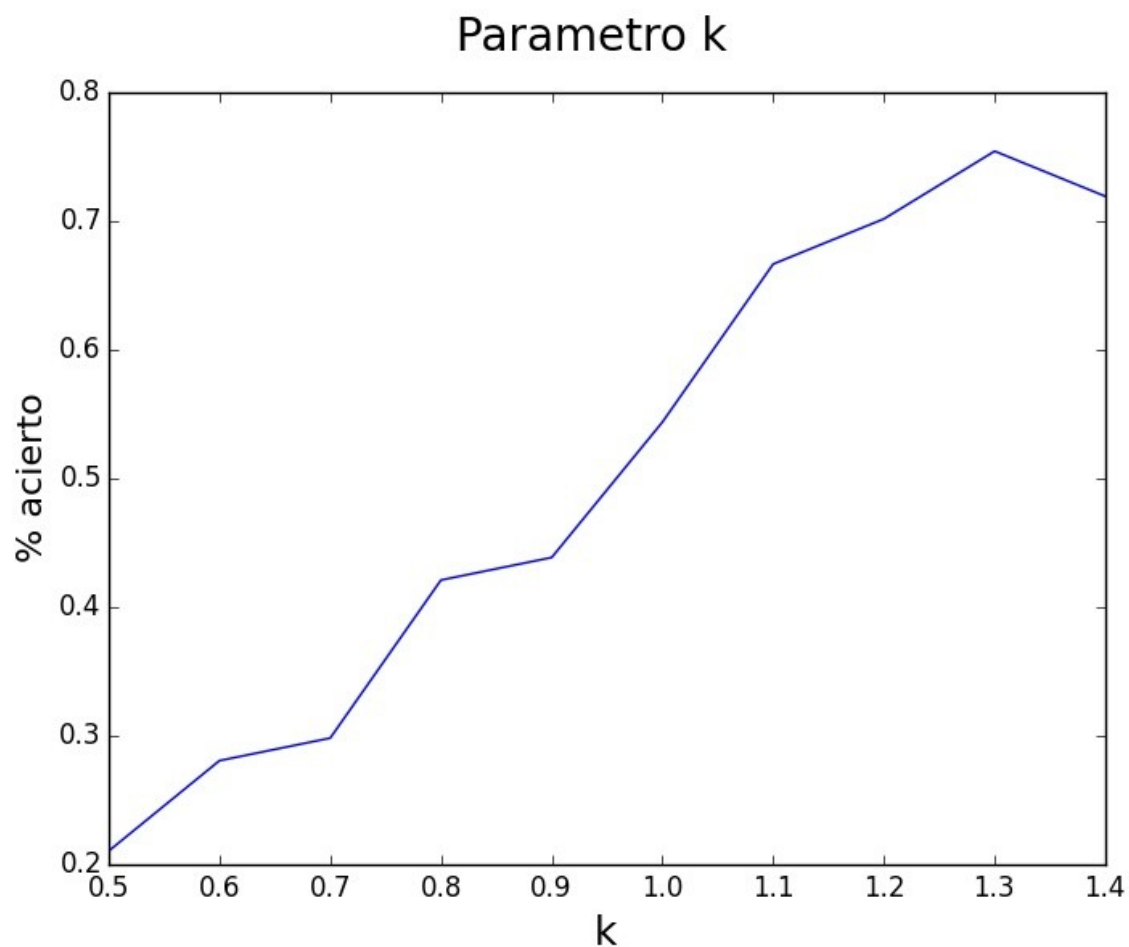
	Umbral inferior*	Umbral Superior**
Parámetro 1	435.39293997	441.58302954
Parámetro 2	18.96876731	19.02318038

RECONOCIMIENTO DE FORMAS – ABRAHAM CARRERA GROBA

Parámetro 3	4.30280411	4.9989391
Parámetro 4	3.22367026	3.82649075
Parámetro 5	1.53611905	1.57644737
Parámetro 6	1.30114598	1.26642928

	Parámetro 1	Parámetro 2	Parámetro 3	Parámetro 4	Parámetro 5	Parámetro 6	Etiqueta
Regla 1	Medio	Medio	Medio	Medio	Medio	Medio	2
Regla 2	Medio	Bajo	Medio	Medio	Medio	Medio	1
Regla 3	Alto	Medio	Medio	Medio	Medio	Medio	1
Regla 4	Medio	Medio	Medio	Bajo	Bajo	Medio	3
Regla 5	Medio	Medio	Medio	Bajo	Medio	Medio	3
Regla 6	Medio	Medio	Medio	Medio	Medio	Alto	1
Regla 7	Medio	Medio	Bajo	Medio	Medio	Medio	1
Regla 8	Medio	Medio	Medio	Medio	Medio	Alto	2
Regla 9	Medio	Medio	Medio	Medio	Alto	Medio	2
Regla 10	Medio	Medio	Alto	Medio	Medio	Medio	2

c) Cáncer

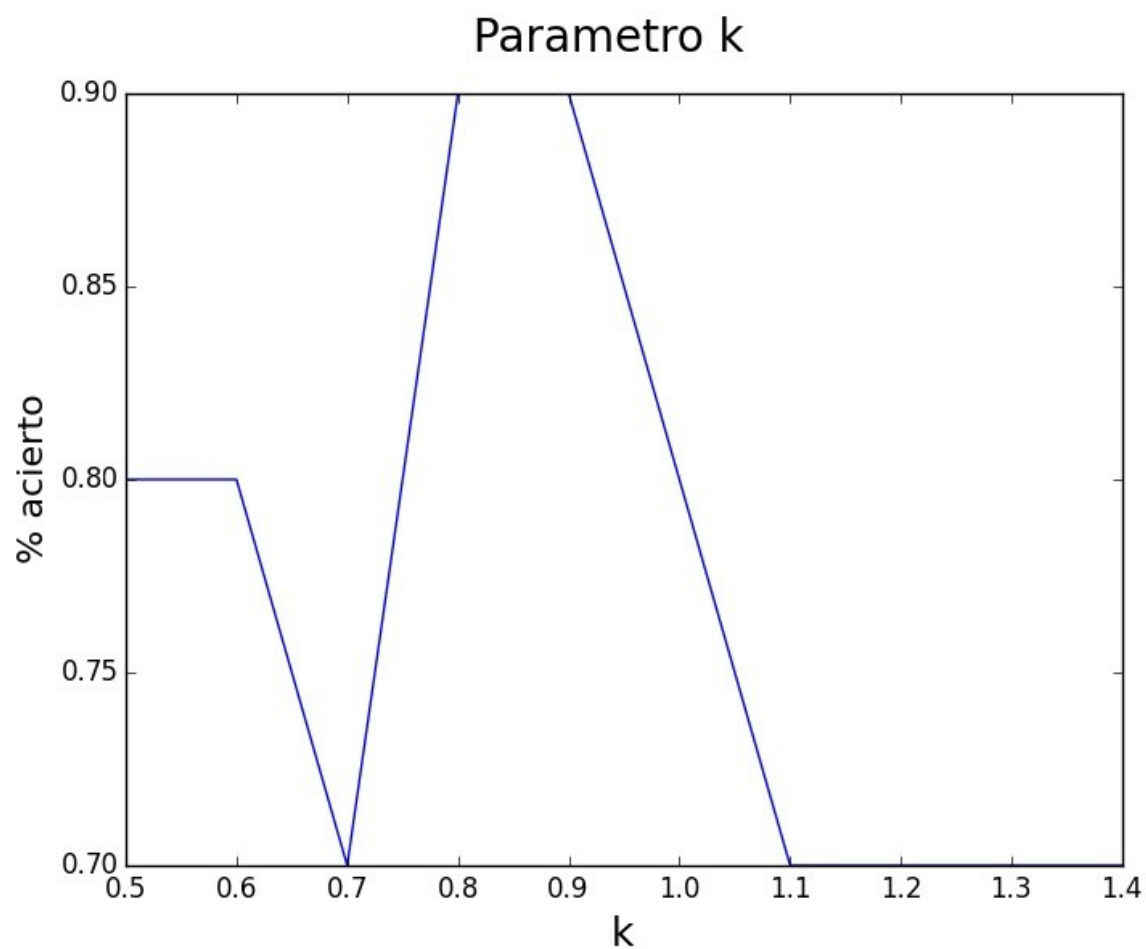


Se alcanza de nuevo el máximo para $k=1.3$.

	Umbral inferior*	Umbral Superior**
Parámetro 1	851.37608009	873.95174415
Parámetro 2	85.81546703	109.04770457
Parámetro 3	14.0820816	34.49543505
Parámetro 4	5.396275	9.4788758
Parámetro 5	5.38486818	8.37328322
Parámetro 6	1.24980477	2.26856347

	Parámetro 1	Parámetro 2	Parámetro 3	Parámetro 4	Parámetro 5	Parámetro 6	Etiqueta
Regla 1	Medio	Medio	Medio	Medio	Medio	Medio	M
Regla 2	Medio	Bajo	Medio	Bajo	Medio	Medio	M
Regla 3	Medio	Medio	Bajo	Medio	Medio	Medio	M
Regla 4	Medio	Medio	Medio	Medio	Alto	Medio	M
Regla 5	Medio	Medio	Medio	Alto	Medio	Medio	B
Regla 6	Alto	Medio	Medio	Medio	Medio	Medio	B
Regla 7	Medio	Medio	Medio	Medio	Medio	Alto	B
Regla 8	Medio	Medio	Medio	Medio	Bajo	Medio	M
Regla 9	Medio	Medio	Medio	Medio	Medio	Bajo	B
Regla 10	Medio	Medio	Medio	Bajo	Bajo	Medio	M

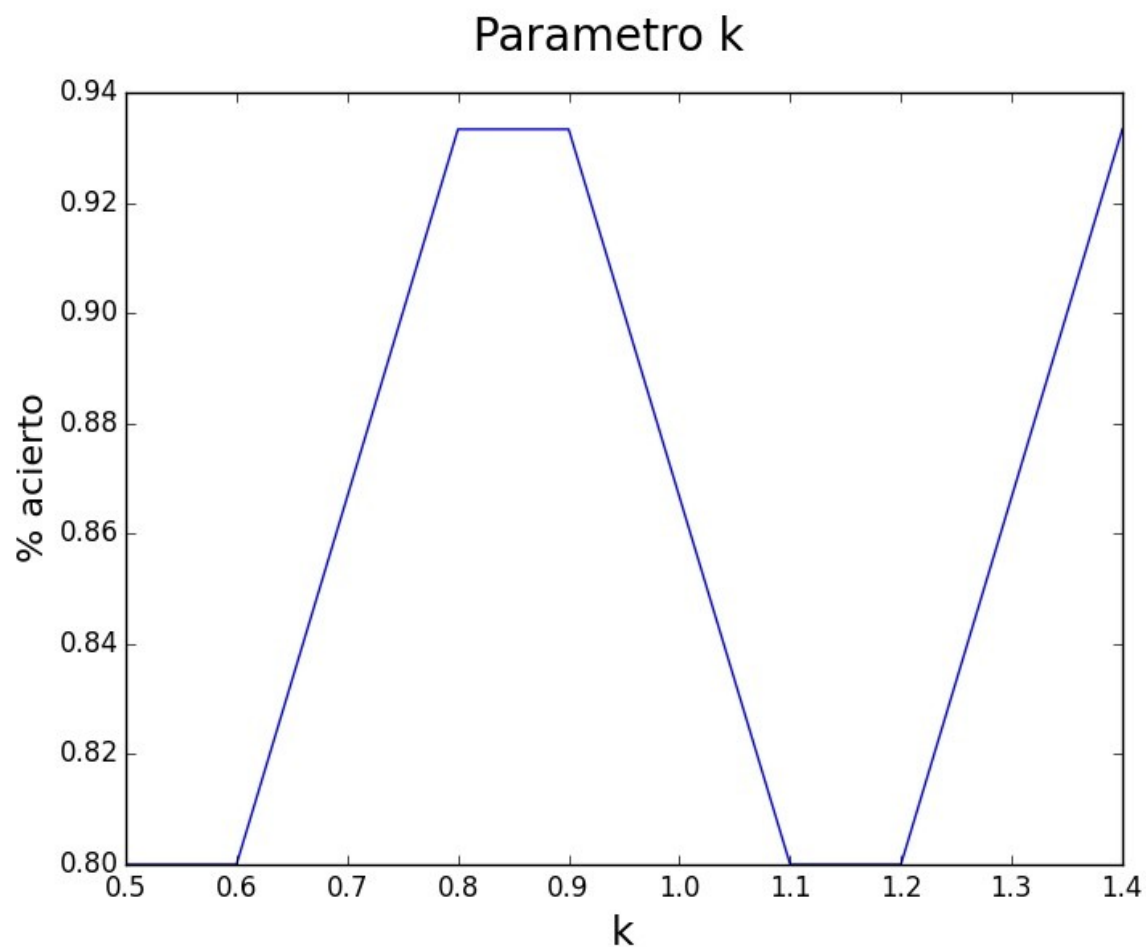
d) Isolet



Se alcanza un máximo de aciertos para $k=0,8$.

	Parámetro 1	Parámetro 2	Parámetro 3	Parámetro 4	Parámetro 5	Parámetro 6	Etiqueta
Regla 1	Alto	Bajo	Medio	Medio	Alto	Medio	7
Regla 2	Bajo	Medio	Bajo	Medio	Medio	Medio	14
Regla 3	Bajo	Medio	Medio	Medio	Medio	Medio	9
Regla 4	Medio	Bajo	Medio	Medio	Alto	Medio	20
Regla 5	Medio	Bajo	Medio	Medio	Medio	Alto	11
Regla 6	Medio	Medio	Alto	Alto	Alto	Medio	6
Regla 7	Alto	Bajo	Medio	Alto	Bajo	Medio	17
Regla 8	Medio	Medio	Medio	Medio	Bajo	Medio	8
Regla 9	Bajo	Bajo	Medio	Bajo	Medio	Medio	2
Regla 10	Bajo	Medio	Bajo	Bajo	Medio	Bajo	1

e) MNIST



Acierto máximo para $k=0,8$.

	Parámetro 1	Parámetro 2	Parámetro 3	Parámetro 4	Parámetro 5	Parámetro 6	Etiqueta
Regla 1	Medio	Alto	Medio	Medio	Medio	Medio	9
Regla 2	Bajo	Medio	Medio	Medio	Medio	Alto	1
Regla 3	Medio	Medio	Bajo	Medio	Medio	Medio	3
Regla 4	Medio	Medio	Medio	Bajo	Alto	Medio	8
Regla 5	Medio	Alto	Medio	Alto	Medio	Alto	4
Regla 6	Bajo	Medio	Medio	Medio	Bajo	Medio	1
Regla 7	Bajo	Medio	Medio	Medio	Medio	Medio	5
Regla 8	Bajo	Medio	Medio	Medio	Medio	Medio	9
Regla 9	Medio	Bajo	Bajo	Medio	Medio	Medio	3
Regla 10	Medio	Bajo	Medio	Bajo	Medio	Medio	7

