

Fundamentos de Aprendizaje Automático

Práctica 4

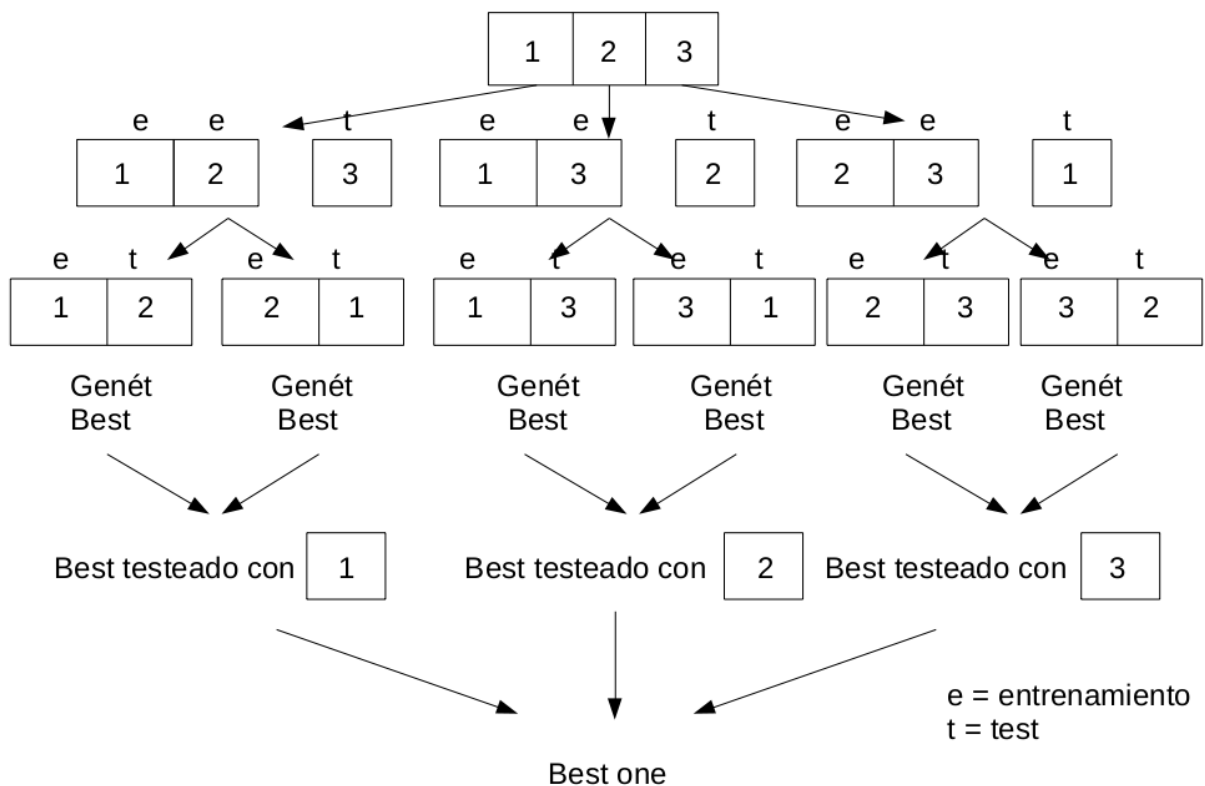
Grupo FAA-1462
Adrián Lorenzo Mateo
Andrés Ruíz Carrasco

Table of Contents

1. Consideraciones respecto al diseño.....	3
2. Ejecución de las aplicaciones.....	4
2.1 Optimización Perceptrón.....	4
2.2 Optimización IBk.....	4
2.3 Ensemble App.....	4
3. Resultados.....	5
3.1 Resultados optimización.....	5
3.1.1 Resultados optimización Perceptrón Multicapa.....	5
3.1.1.1 Resultados sobre credit-g.....	5
3.1.1.2 Resultados sobre segment-test.....	7
3.1.1.3 Resultados sobre breast-cancer.....	8
3.1.1.4 Resultados sobre glass.....	10
3.1.2 Resultados optimización IBk.....	12
3.1.2.1 Resultados sobre credit-g.....	12
3.1.2.2 Resultados sobre segment-test.....	13
3.1.2.3 Resultados sobre breast-cancer.....	15
3.1.2.4 Resultados sobre glass.....	16
3.1.3 Análisis de los resultados.....	18
3.2 Resultados ensemble vs clasificadores individuales.....	19
3.2.1 Análisis de los resultados Ensemble vs Clasificadores.....	21
Anexo: Información de los conjuntos de datos.....	22
1. Conjunto Credit-g.....	22
Distribución clases.....	22
Distribución atributo checkingStatus.....	23
2. Conjunto Segment-test.....	24
3. Conjunto Breast-cancer.....	25
4. Conjunto Glass.....	27

1. Consideraciones respecto al diseño

La optimización se va a realizar a través de la aplicación de un algoritmo genético. Cada individuo será una red neuronal o un clasificador kNN. Las mutaciones consistirán en cambiar el número de vecinos próximos o el número de neuronas de capa oculta o la constante de aprendizaje. Del mismo modo ocurre en la fase de cruce. El funcionamiento del algoritmo que explicado en la siguiente ilustración:



2. Ejecución de las aplicaciones

2.1 Optimización Perceptrón

- -input <fichero>: Ruta del fichero .arff (contenido en la carpeta data).
- -nFolds <numero>: Número de folds para el método de validación cruzada.
- -epochs <numero>: Número de épocas del algoritmo genético.
- -population <numero>: Número de individuos en la población.

2.2 Optimización IBk

- -input <fichero>: Ruta del fichero .arff (contenido en la carpeta data).
- -nFolds <numero>: Número de folds para el método de validación cruzada.
- -epochs <numero>: Número de épocas del algoritmo genético.
- -population <numero>: Número de individuos en la población.

2.3 Ensemble App

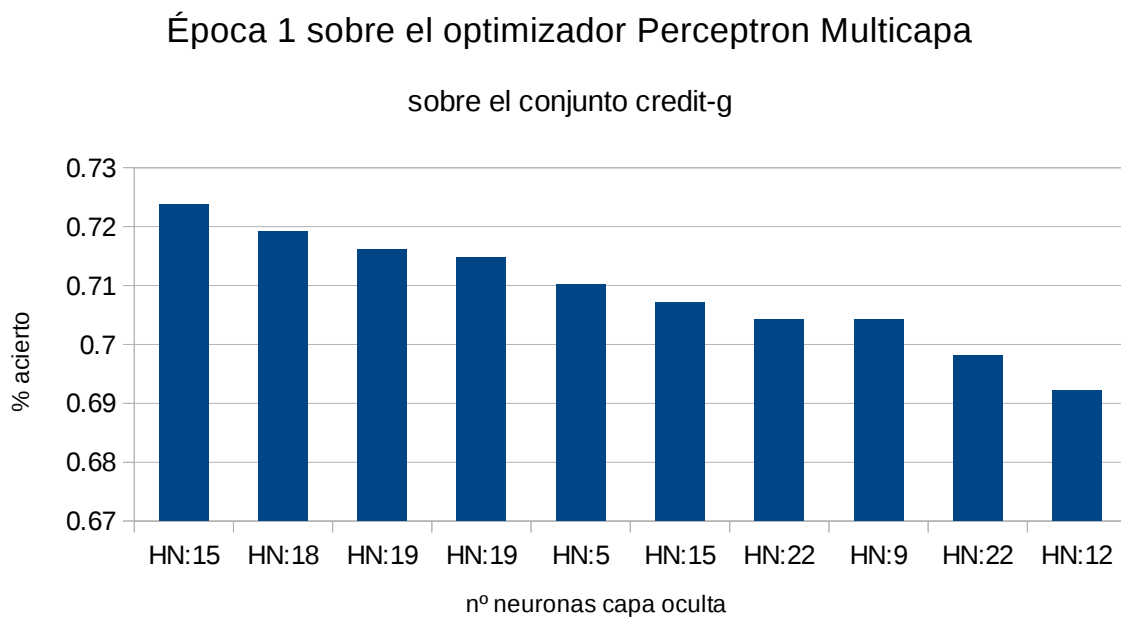
- -input <fichero>: Ruta del fichero .arff (contenido en la carpeta data).
- -nFolds <numero>: Número de folds para el método de validación cruzada.
- -neurons <numero>: Número de neuronas en la (única) capa oculta.
- -K <numero>: Tamaño del vecindario para el clasificador KNN.
- -epochs <numero>: Número de épocas de los clasificadores.

3. Resultados

3.1 Resultados optimización

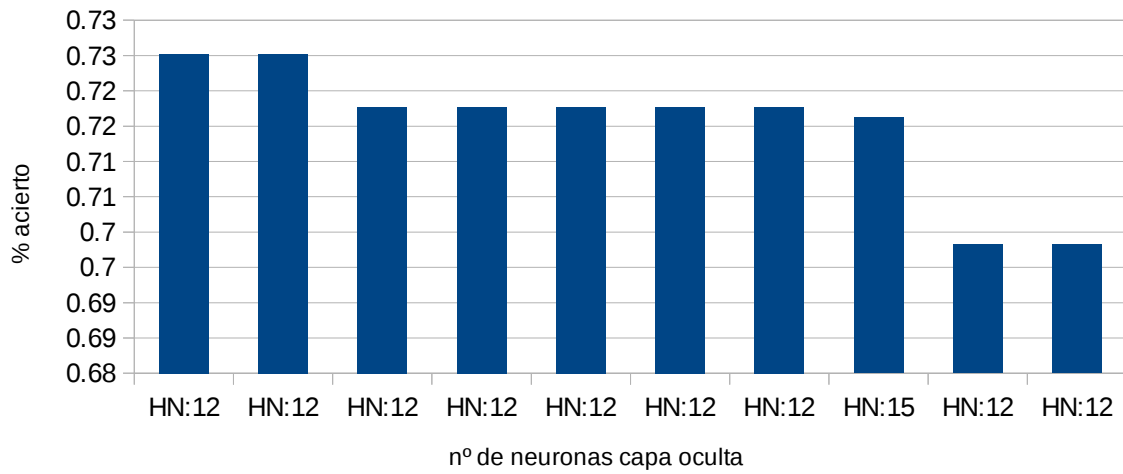
3.1.1 Resultados optimización Perceptrón Multicapa

3.1.1.1 Resultados sobre credit-g



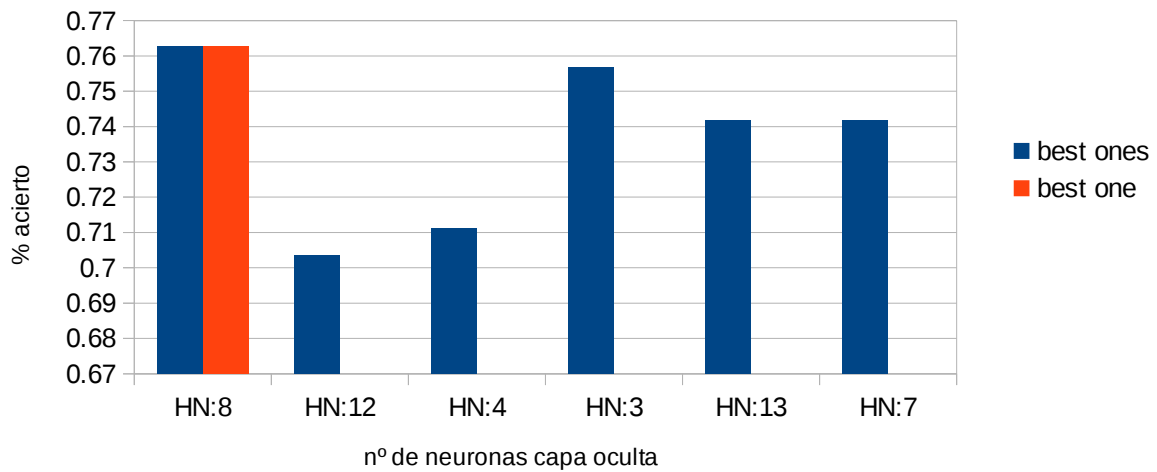
Época 10 sobre el optimizador Perceptron Multicapa

sobre el conjunto credit-g



Fase de test sobre el optimizador Perceptron Multicapa

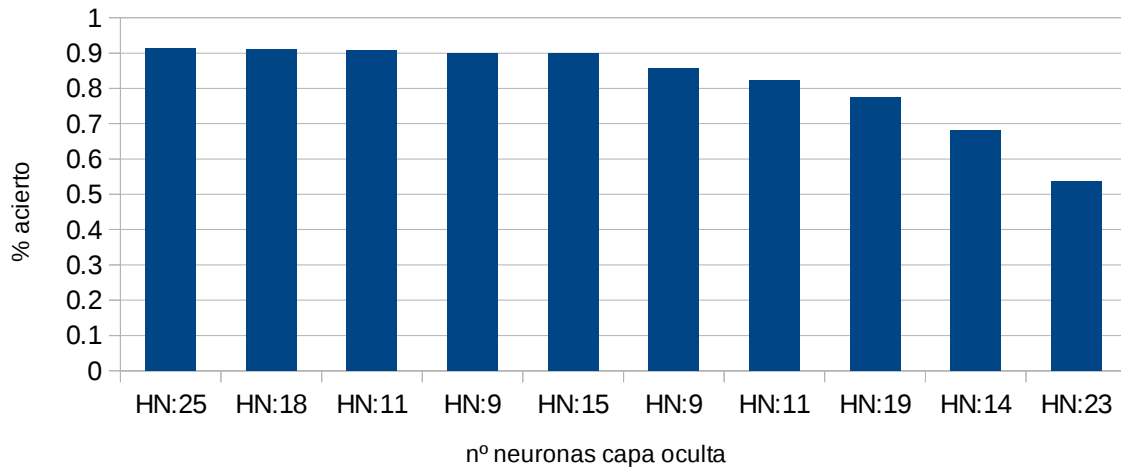
sobre el conjunto credit-g (nFolds = 3)



3.1.1.2 Resultados sobre segment-test

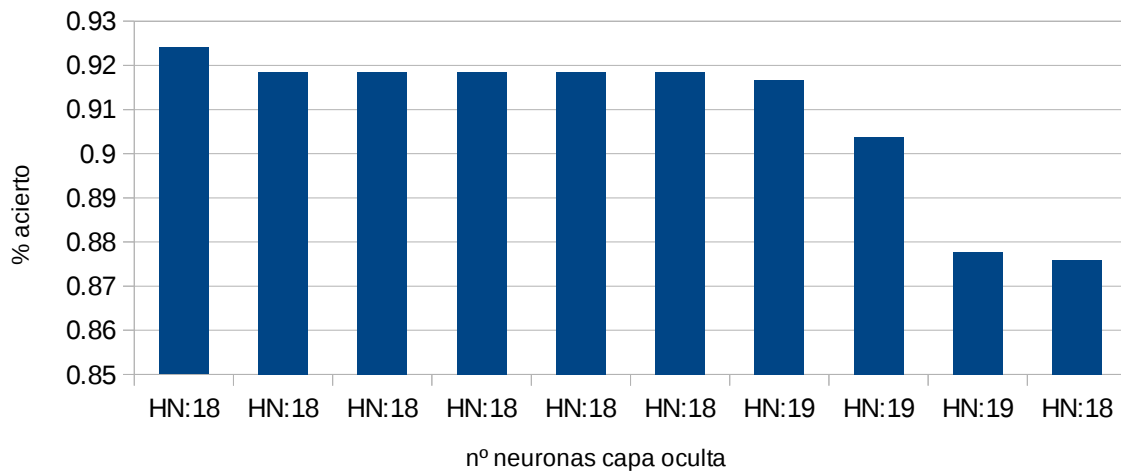
Época 1 sobre el optimizador Perceptron Multicapa

sobre el conjunto segment-test



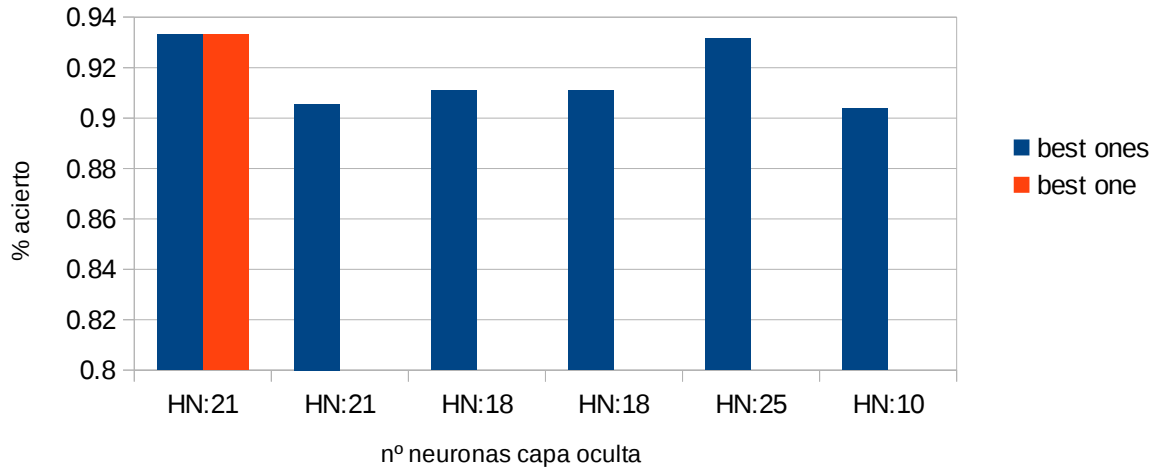
Época 10 sobre el optimizador Perceptron Multicapa

sobre el conjunto segment-test



Fase de test sobre el optimizador Perceptron Multicapa

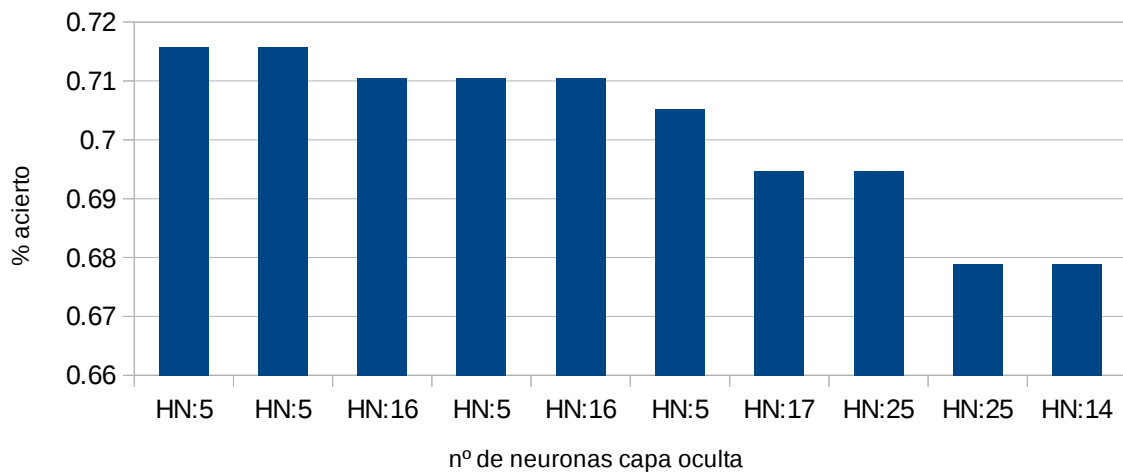
sobre el conjunto segment-test (nFolds=3)



3.1.1.3 Resultados sobre breast-cancer

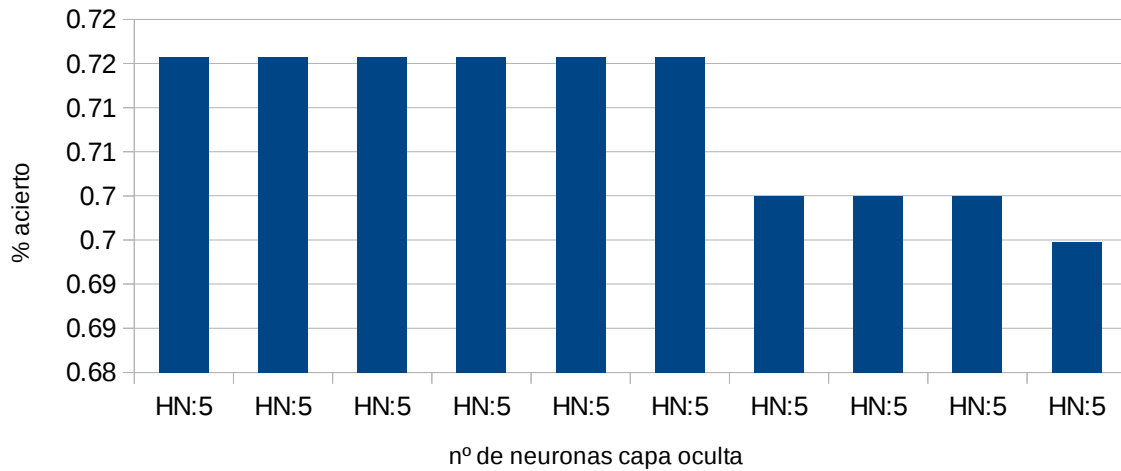
Época 1 sobre el optimizador Perceptrón Multicapa

sobre el conjunto breast-cancer



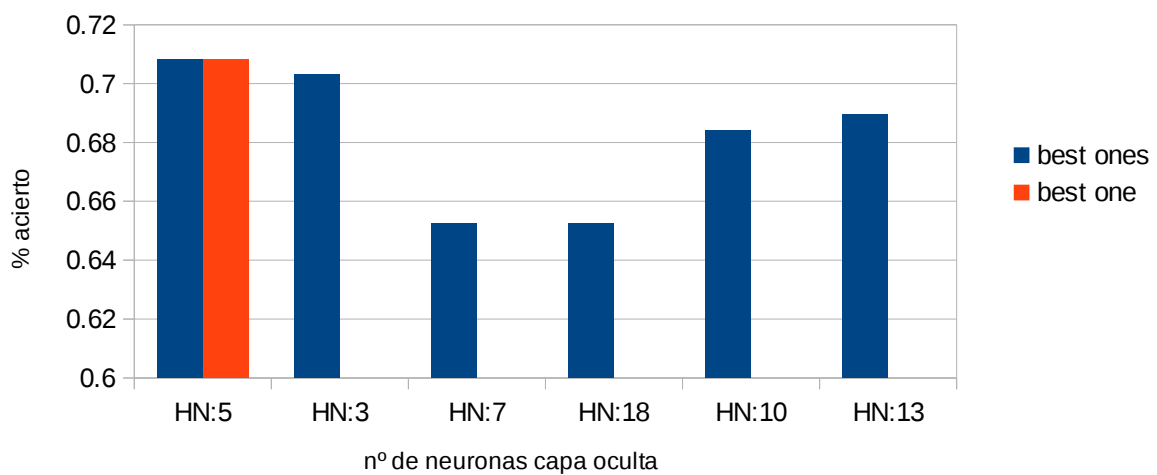
Época 10 sobre el optimizador Perceptron Multicapa

sobre el conjunto breast-cancer



Fase de test sobre el optimizador Perceptron Multicapa

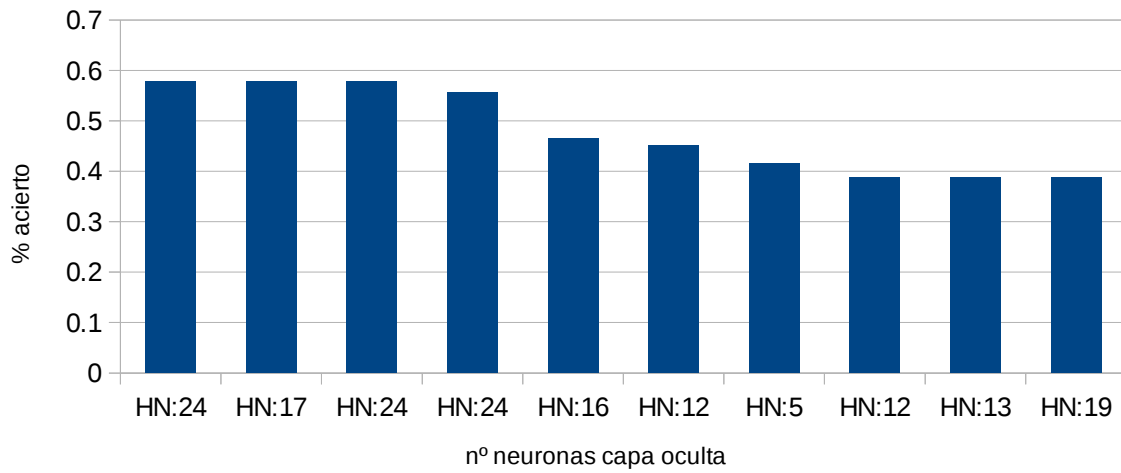
sobre el conjunto breast-cancer (nFolds = 3)



3.1.1.4 Resultados sobre glass

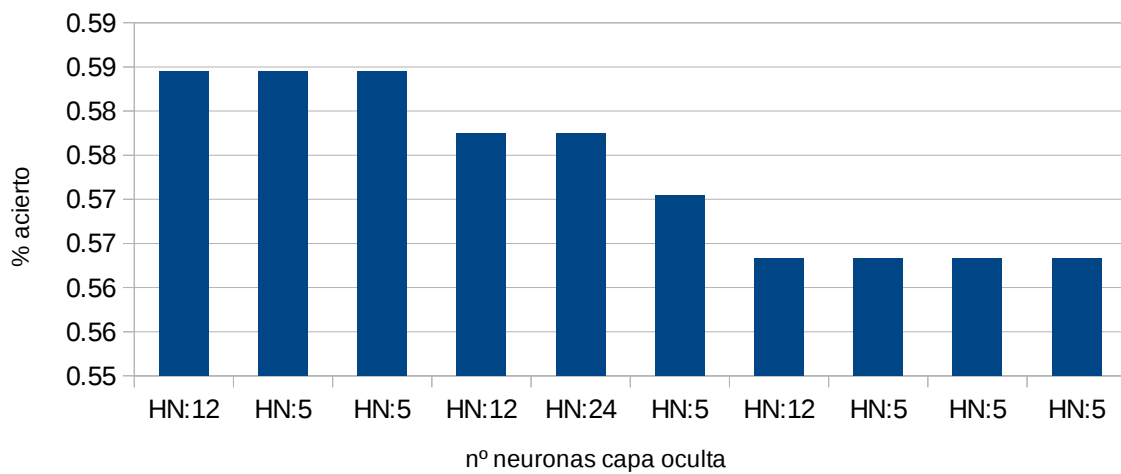
Época 1 sobre el optimizador Perceptron Multicapa

sobre el conjunto glass



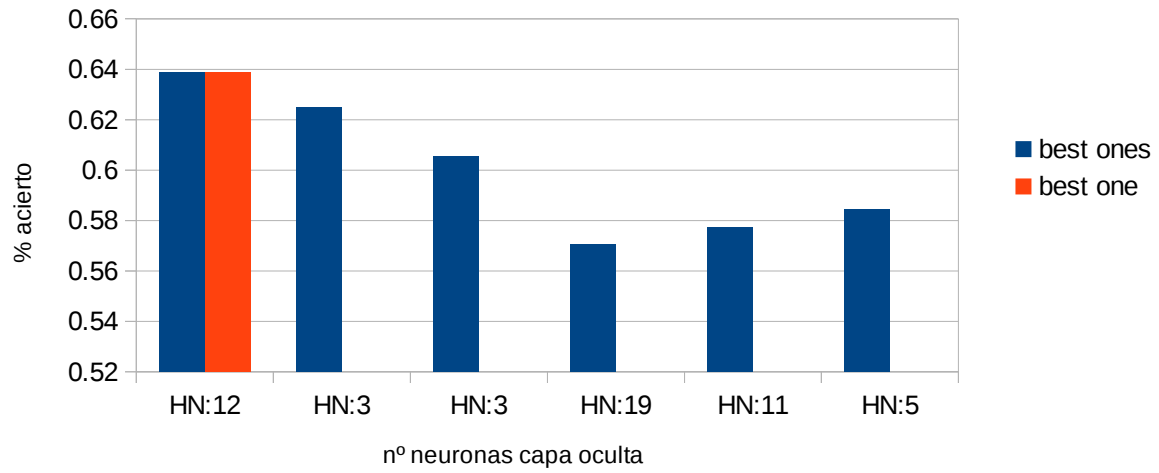
Época 10 sobre el optimizador Perceptron Multicapa

sobre el conjunto glass



Fase de test sobre el optimizador Perceptron Multicapa

sobre el conjunto glass (nFolds = 3)

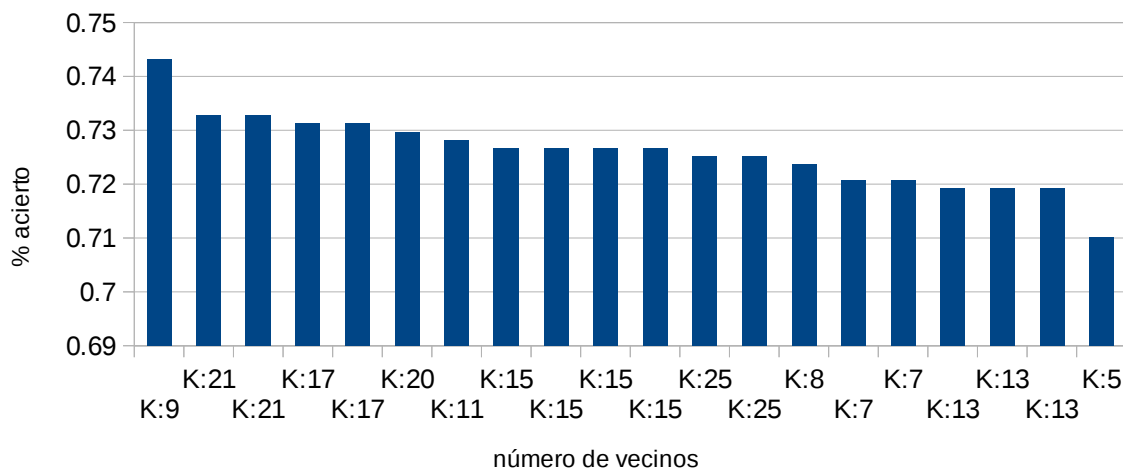


3.1.2 Resultados optimización IBk

3.1.2.1 Resultados sobre credit-g

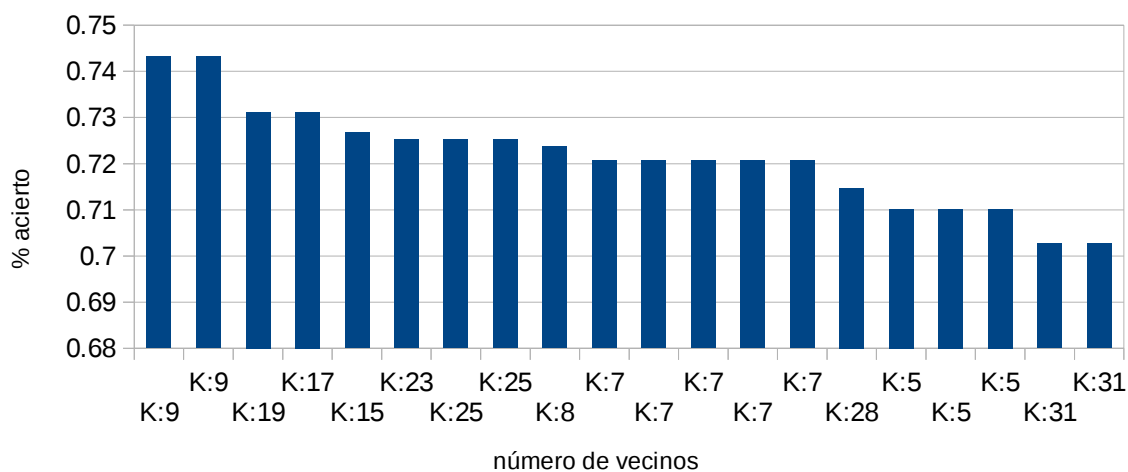
Época 1 de la optimización de IBk

sobre el conjunto de credit-g



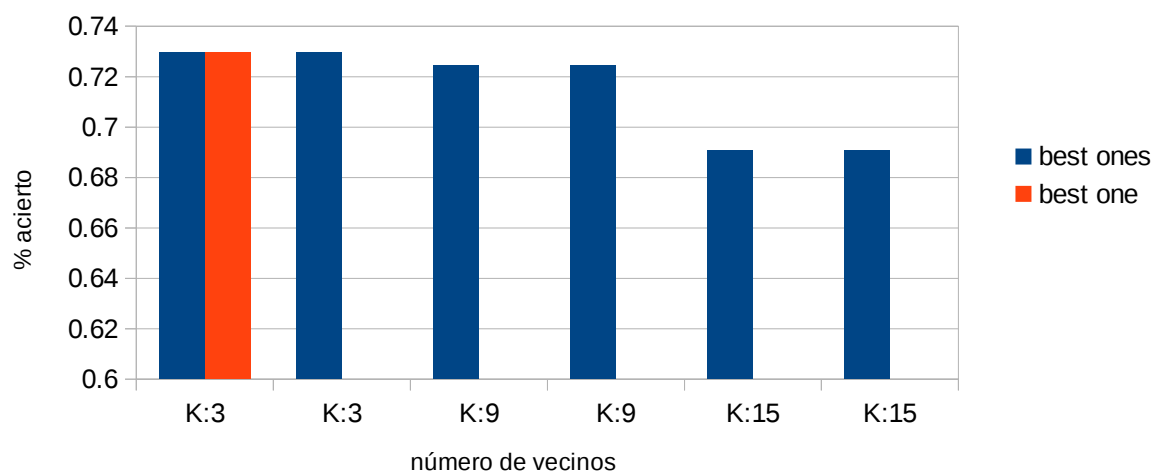
Época 30 de la optimización de IBk

sobre el conjunto credit-g



Fase Test sobre los mejores individuos

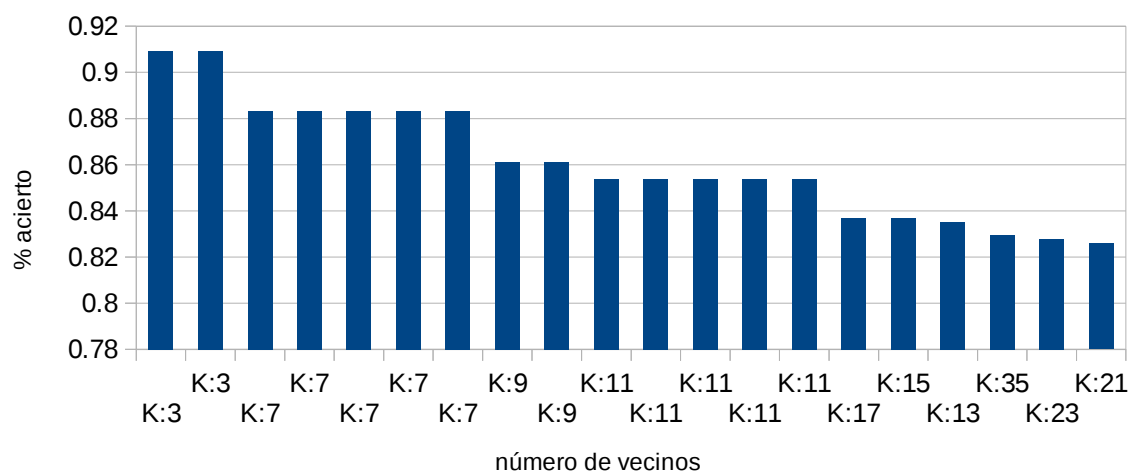
sobre el conjunto credit-g (nFolds = 3)



3.1.2.2 Resultados sobre segment-test

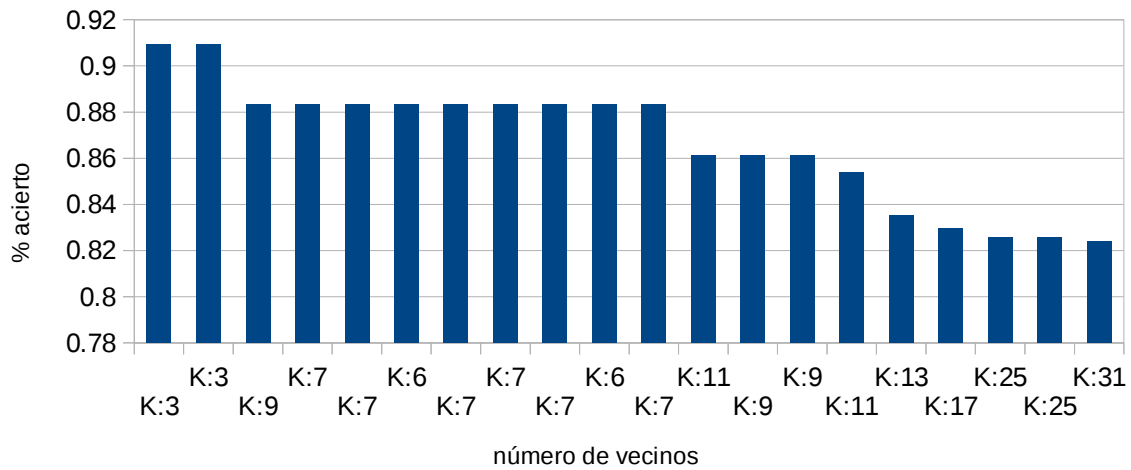
Época 1 de la optimización de IBk

sobre el conjunto segment-test



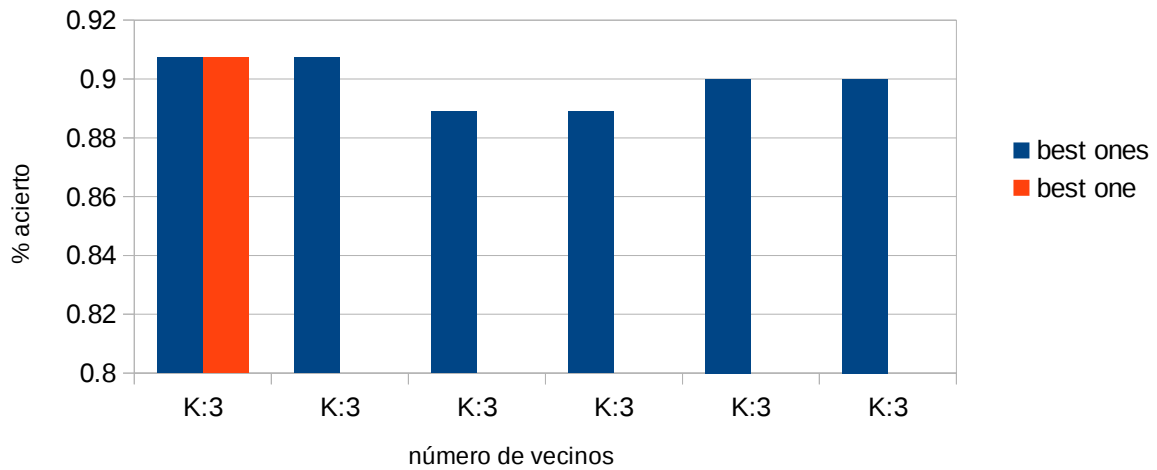
Época 30 de la optimización de IBk

sobre el conjunto segment-test



Fase de test sobre optimizador de IBk

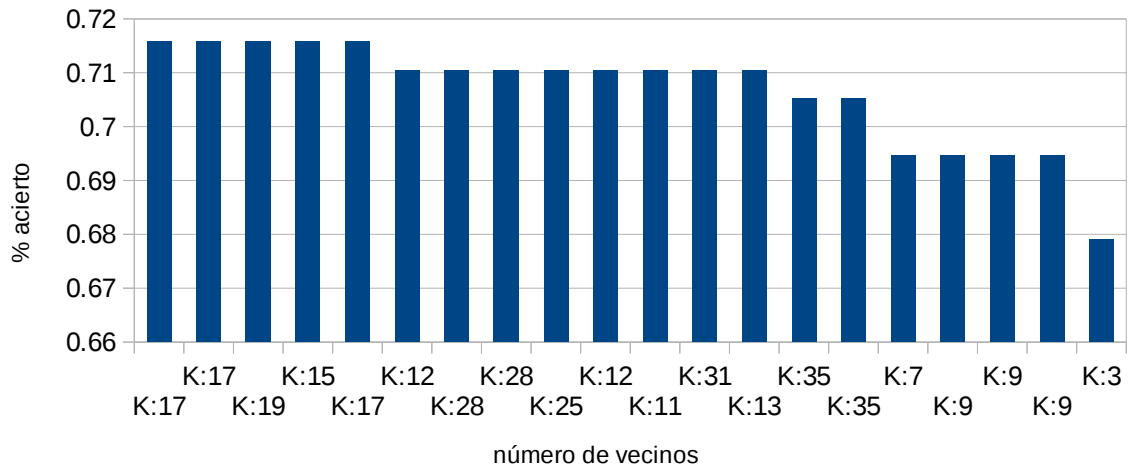
sobre el conjunto de segment-test (nFolds = 3)



3.1.2.3 Resultados sobre breast-cancer

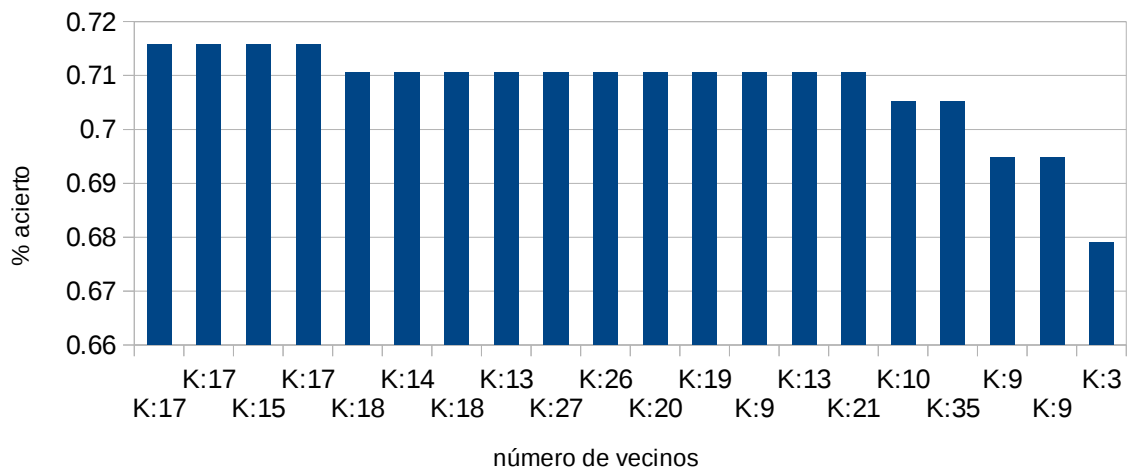
Época 1 sobre el optimizador IBk

sobre el conjunto breast-cancer



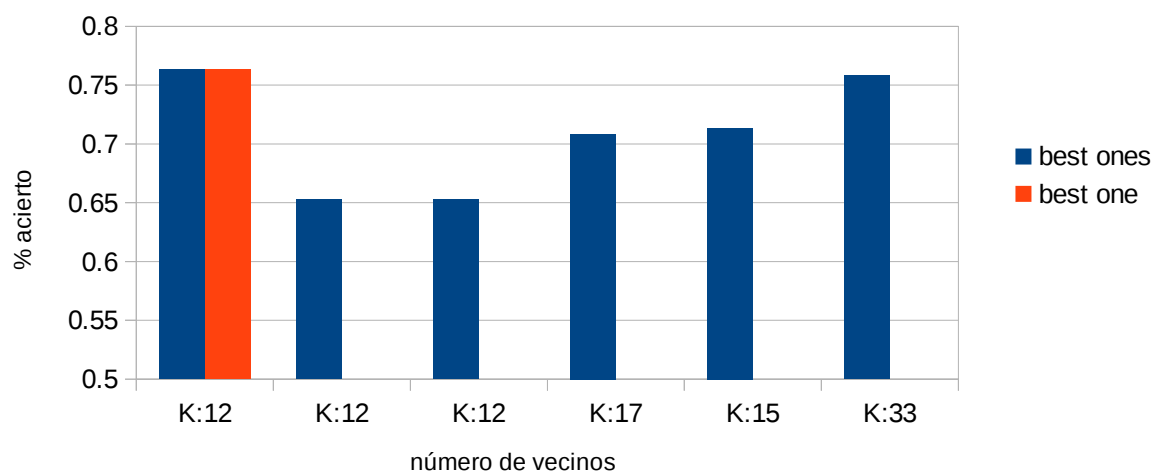
Época 30 sobre el optimizador IBk

sobre el conjunto breast-cancer



Fase de test sobre el optimizador IBk

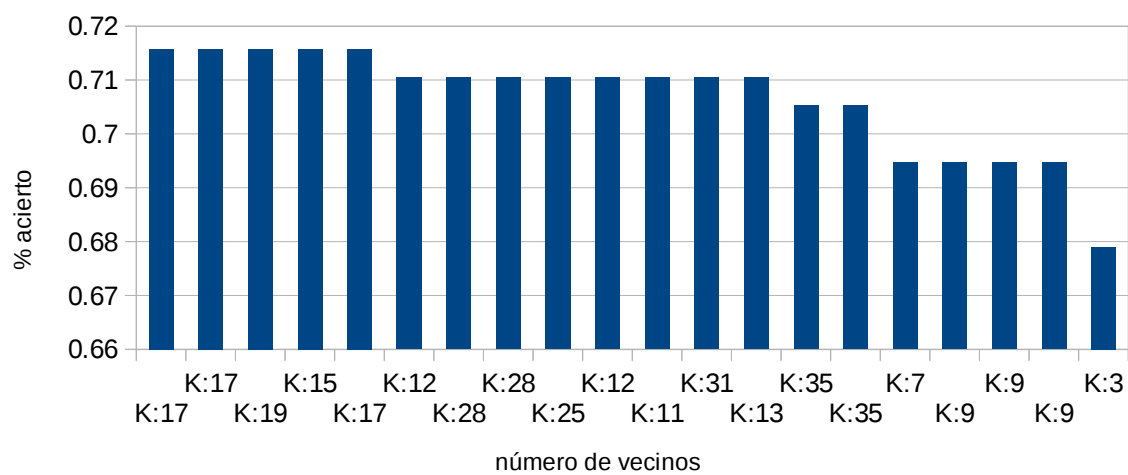
sobre el conjunto breast-cancer (nFolds = 3)



3.1.2.4 Resultados sobre glass

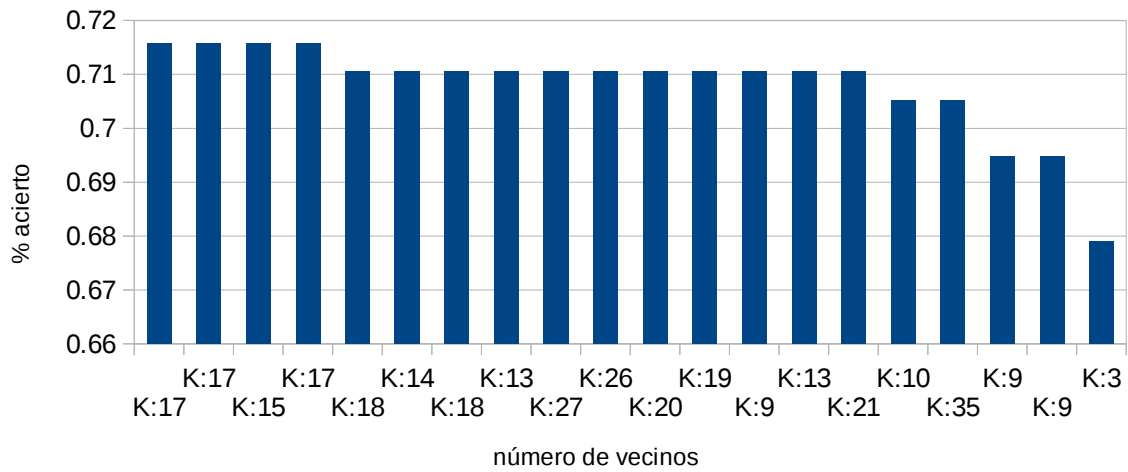
Época 1 sobre el optimizador IBk

sobre el conjunto glass



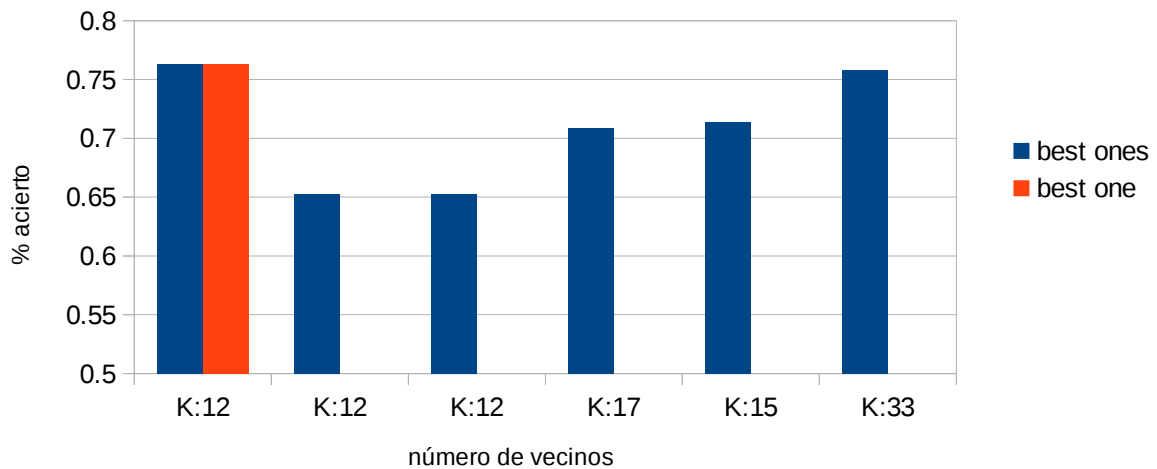
Época 30 sobre el optimizador IBk

sobre el conjunto glass



Fase de test sobre el optimizador IBk

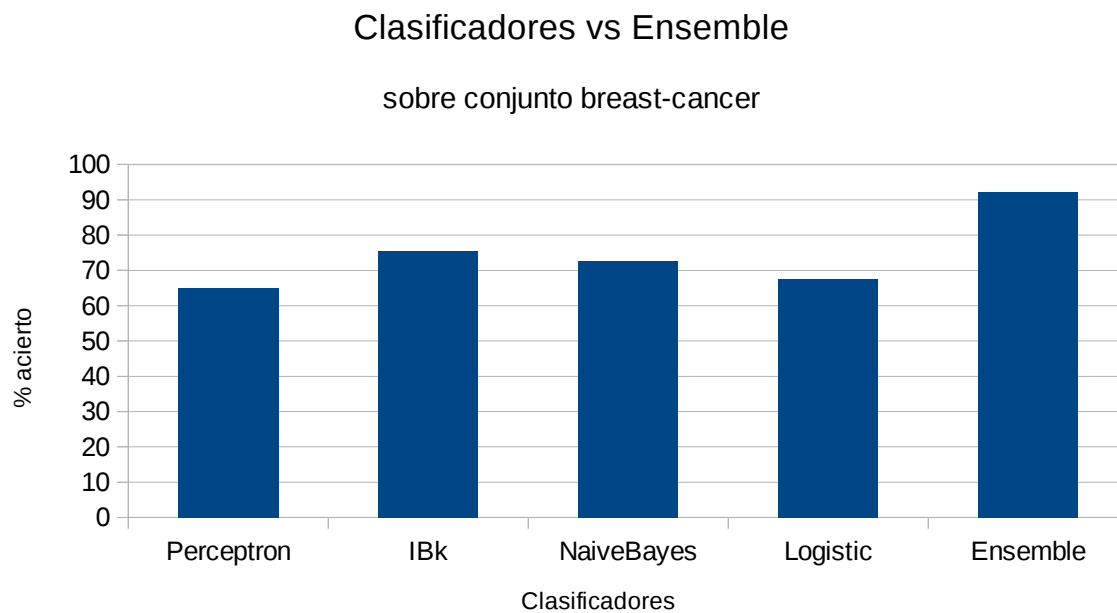
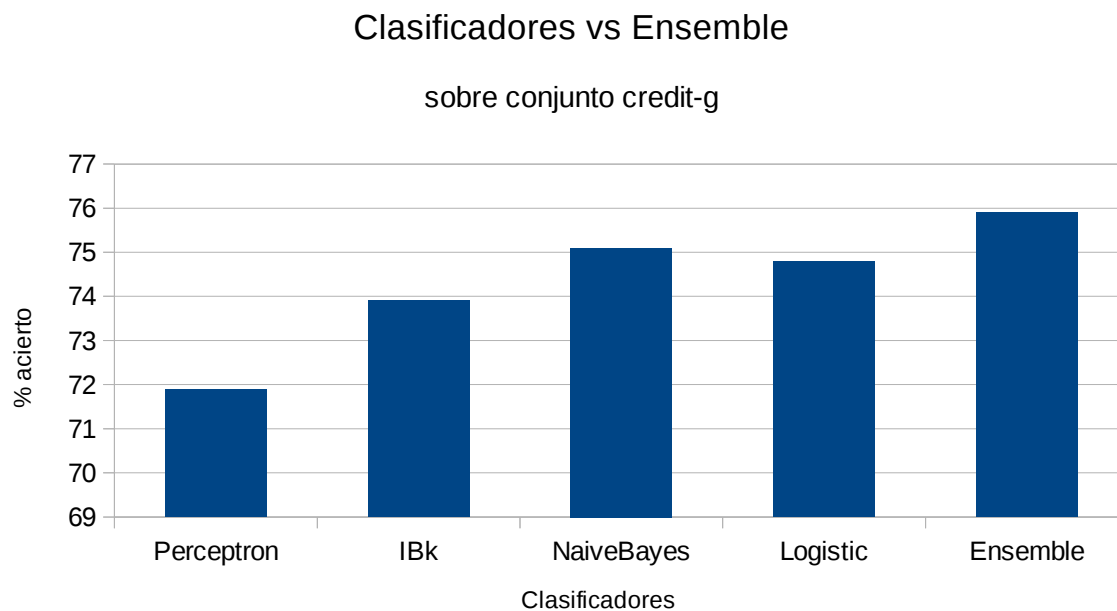
sobre el conjunto glass (nFolds = 3)



3.1.3 Análisis de los resultados

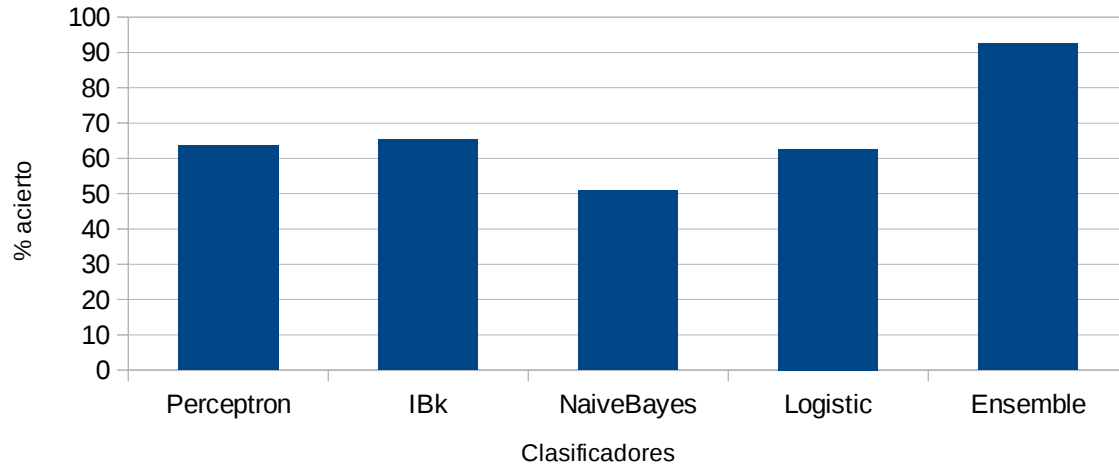
Por lo general los resultados para el perceptrón multicapa y para kNN son parecidos. En la mayoría de los casos durante la primera época del algoritmo genético ya se obtienen las configuraciones óptimas, explorando en las siguientes etapas la posibilidad de que existan mejores configuraciones. La gran mayor parte de resultados se posicionan por encima del 70% de acierto excepto en el caso del perceptrón sobre el conjunto glass, lo que puede ser ocasionado por el número de épocas (1000), que posiblemente sea insuficiente para la convergencia de la red neuronal para este conjunto de datos.

3.2 Resultados ensemble vs clasificadores individuales



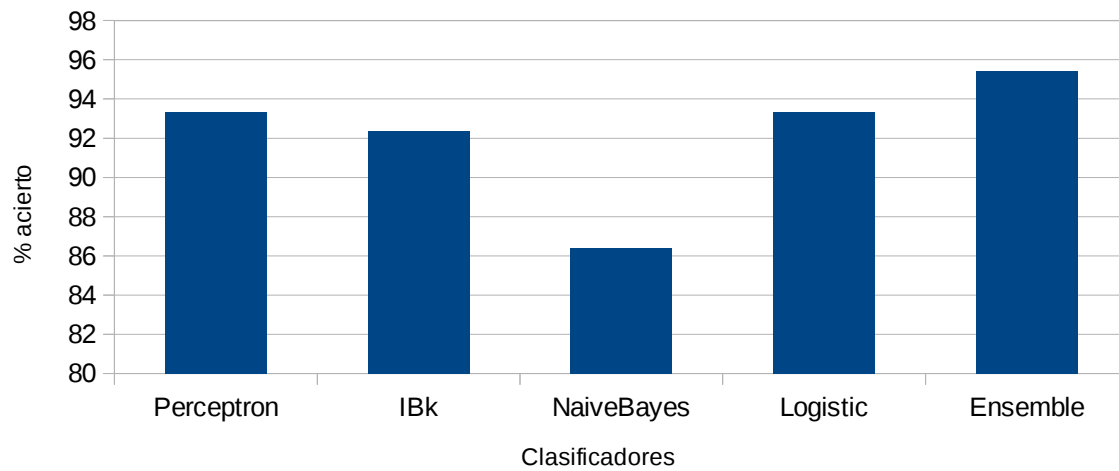
Clasificadores vs Ensemble

sobre conjunto glass



Clasificadores vs Ensemble

sobre conjunto segment-test



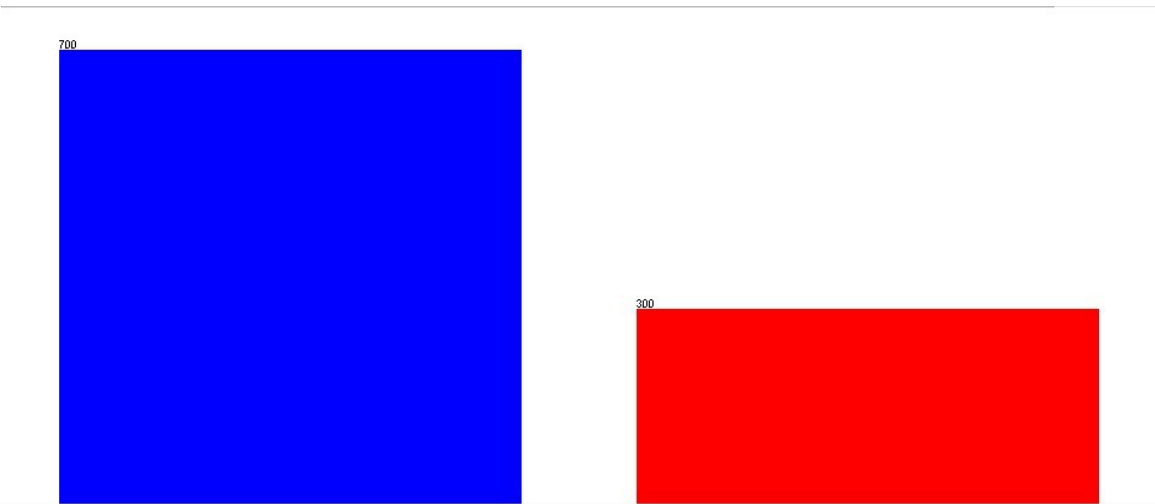
3.2.1 Análisis de los resultados Ensemble vs Clasificadores

Como se puede comprobar en las gráficas, el método Ensemble de clasificación da mejores resultados que los clasificadores individuales. El método Ensemble que se ha escogido es el voto mayoritario, lo que hace que los errores de los clasificadores individuales sean eliminados o suavizados por el promedio. Por lo general la capacidad de generalización de los clasificadores Ensemble es superior de los individuales ya que a veces el conjunto de entrenamiento no proporciona suficiente información para realizar una clasificación, el proceso de búsqueda de los algoritmos de aprendizaje automático puede ser imperfecto o el espacio de hipótesis sobre el que se busca puede no contener la verdadero función objetivo, mientras que los clasificadores Ensemble pueden proporcionar una buena aproximación.

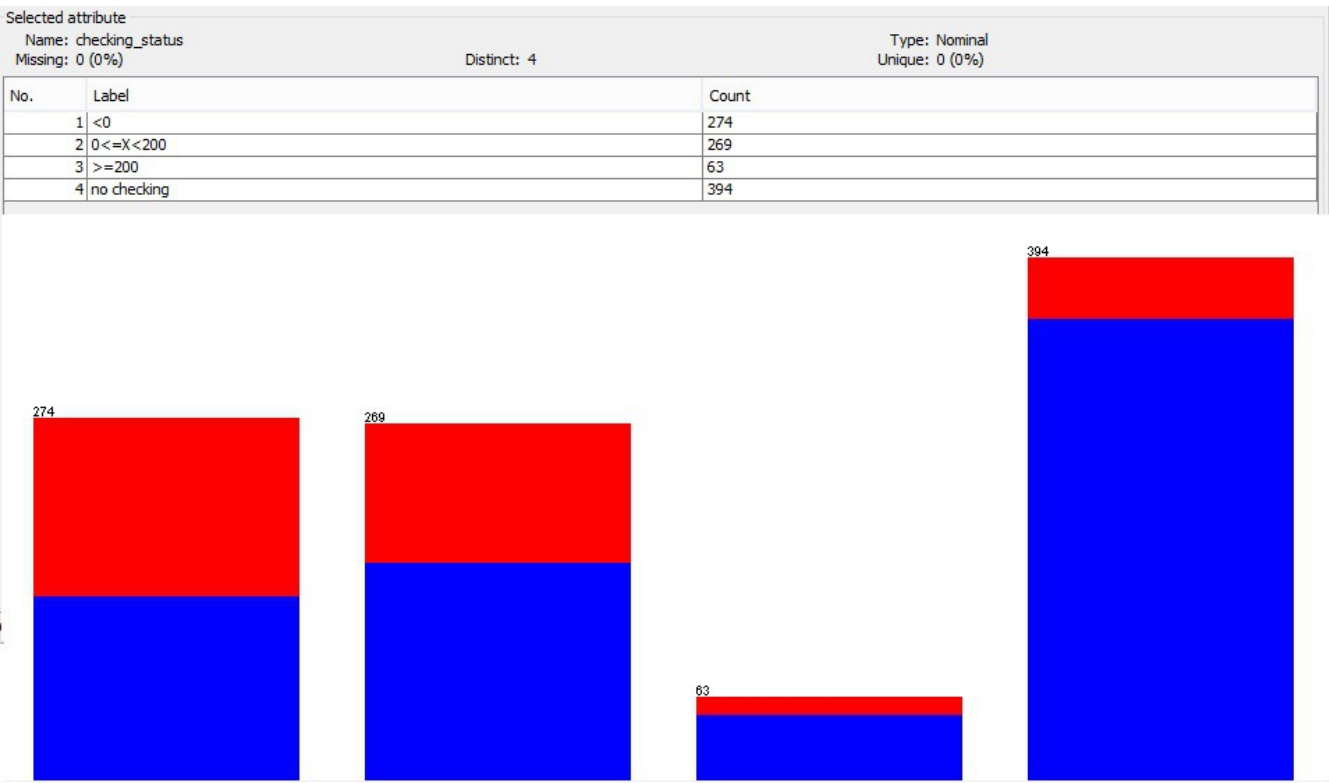
Anexo: Información de los conjuntos de datos

1. Conjunto Credit-g

Distribución clases

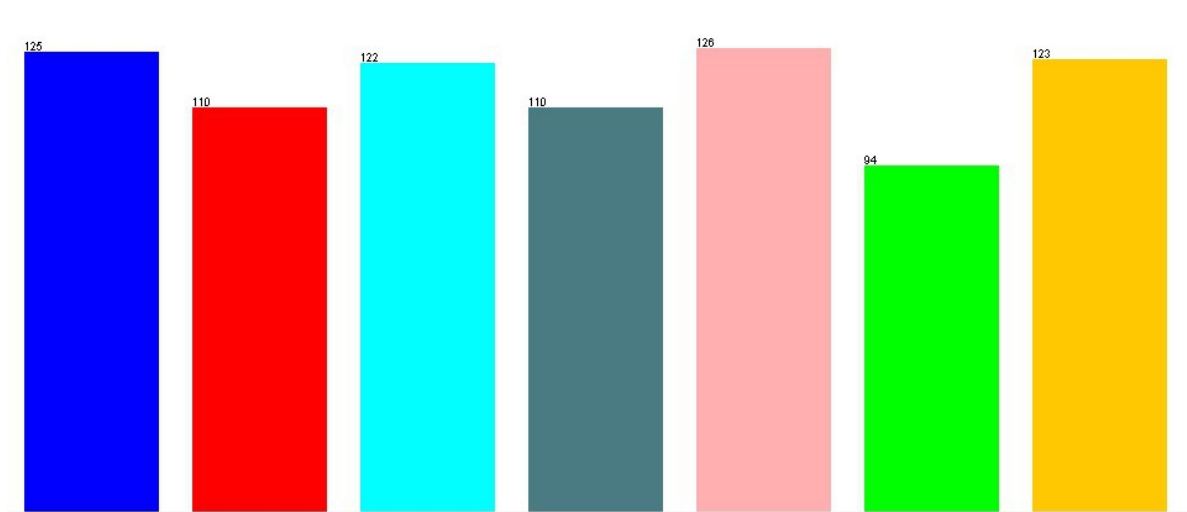


Distribución atributo checkingStatus

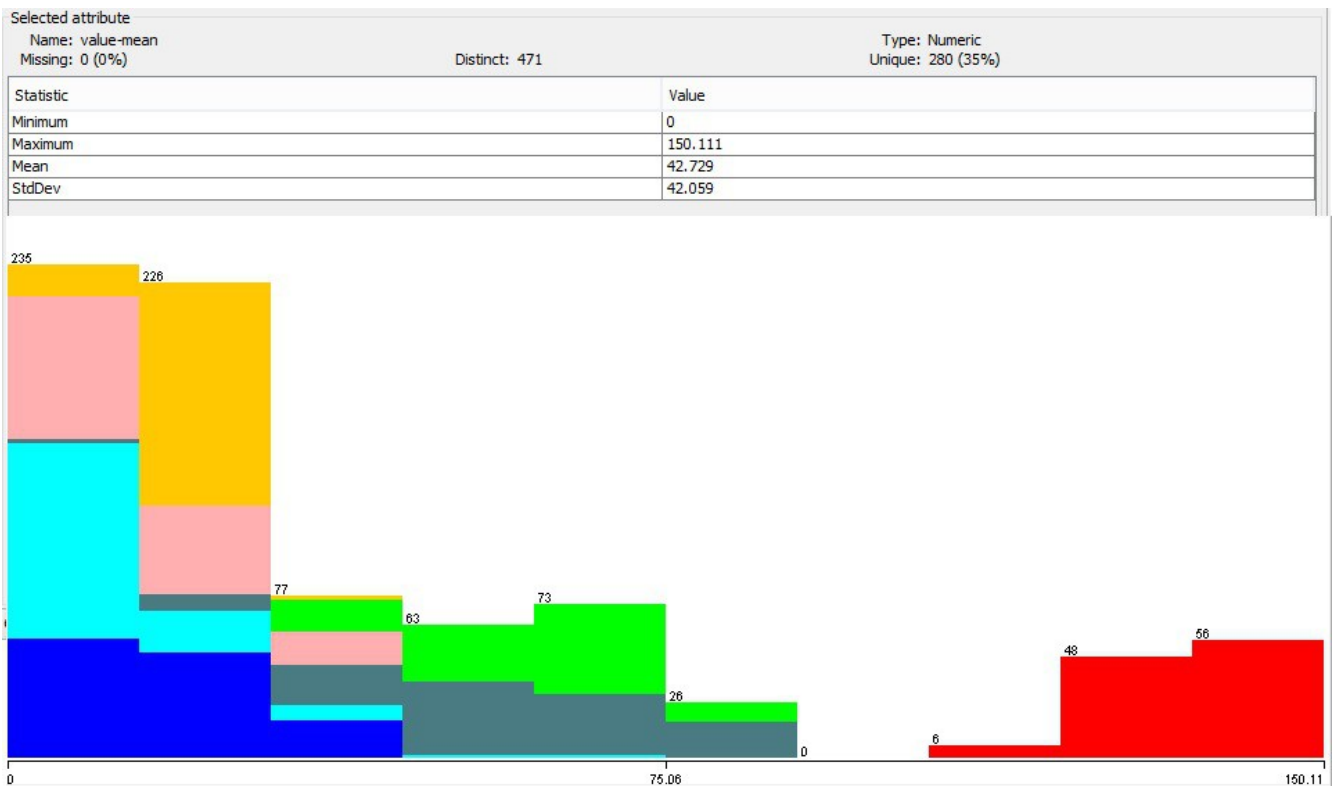


2. Conjunto Segment-test

Distribución clases

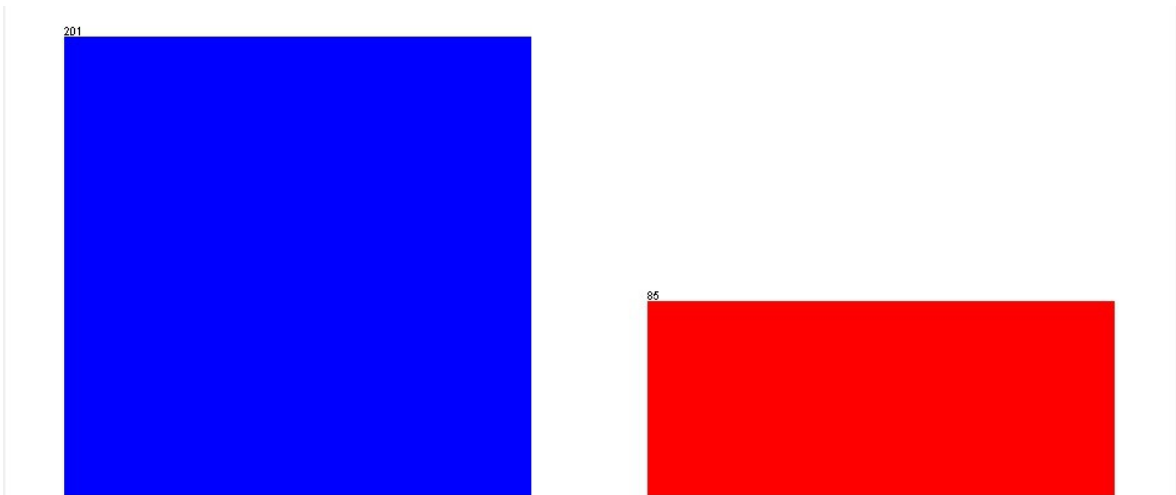


Distribución atributo value-mean

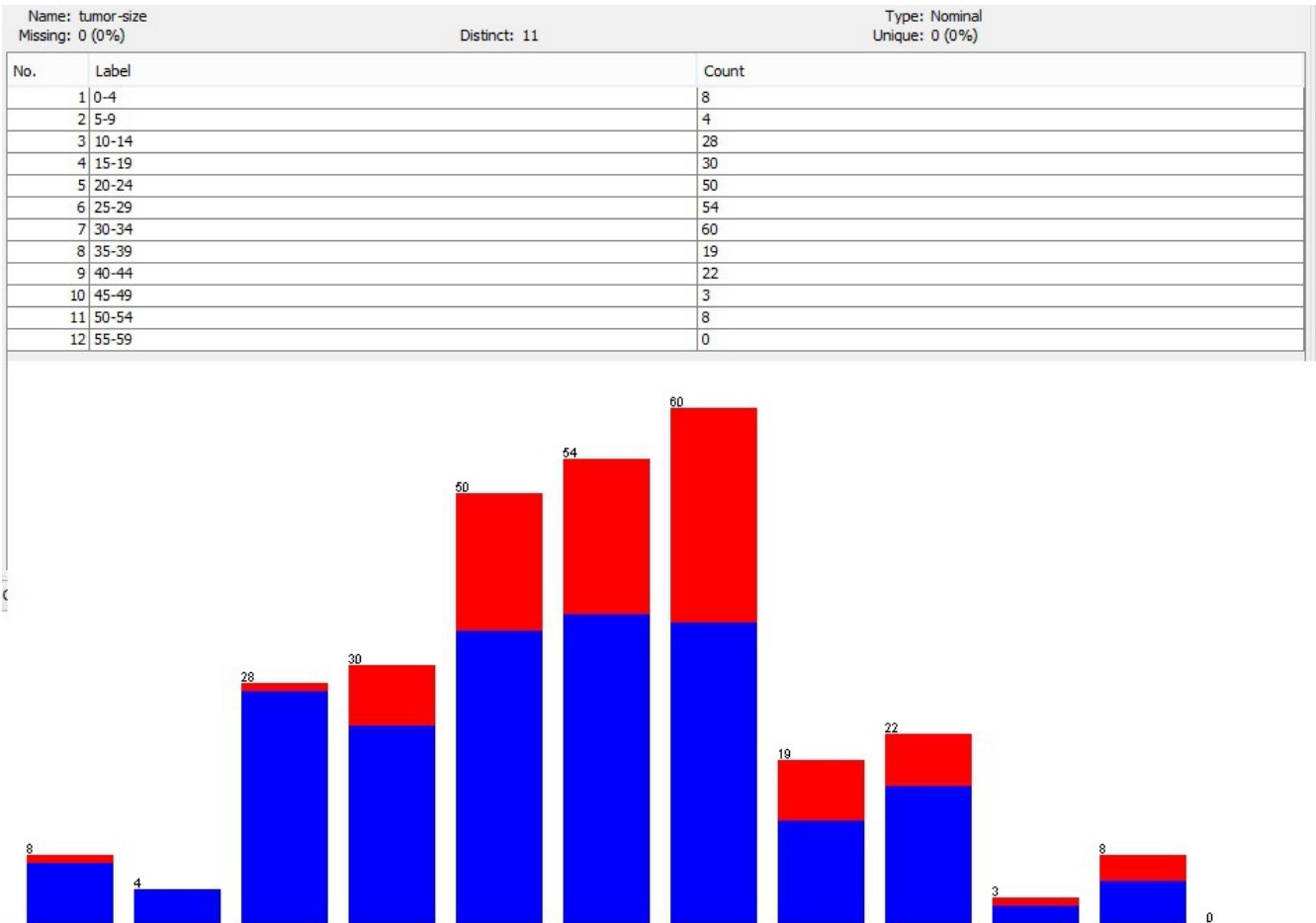


3. Conjunto Breast-cancer

Distribución clases

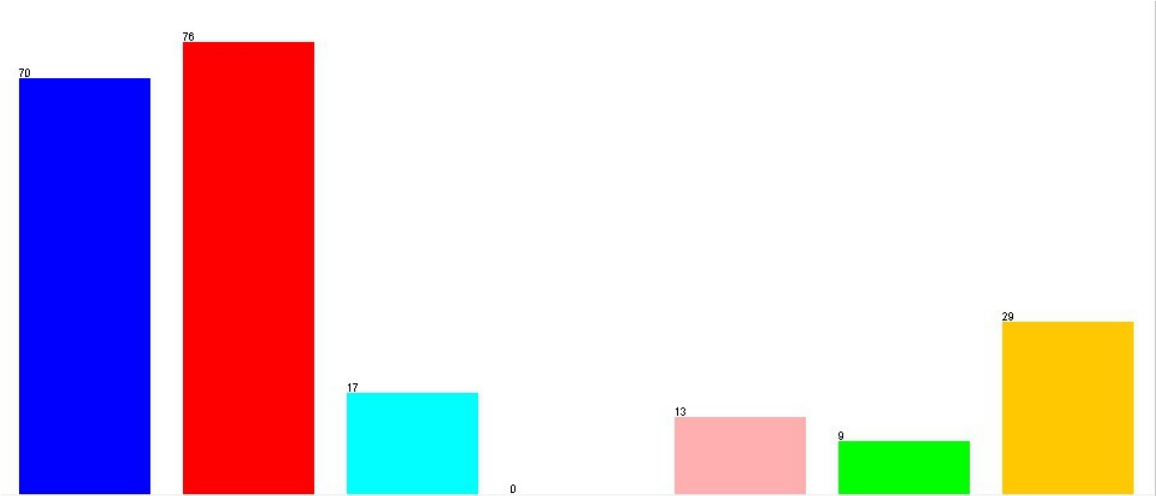


Distribución atributo tumor-size



4. Conjunto Glass

Distribución clases



Distribución atributo Mg

