



UNIVERSIDAD DE CÓRDOBA
ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA

INGENIERÍA INFORMÁTICA
ESPECIALIDAD: COMPUTACIÓN
CUARTO CURSO. PRIMER CUATRIMESTRE

INTRODUCCIÓN A LA MINERÍA DE DATOS

Práctica 2: Clasificación y evaluación de modelos

Valentín Gabriel Avram Aenachioei
03524931C
p92avavv@uco.es

Curso académico 2022-2023
Córdoba, 9 de junio de 2023

Índice

Índice de figuras	III
Índice de tablas	IV
1. Preguntas práctica 2.1	1
1.1. Seleccione como método para obtener el error la validación cruzada de 10 particiones o el método hold out. Seleccione dos métricas de las estudiadas en teoría. Entrene cada clasificador seleccionado y anote el valor de las métricas estudiadas. Fije los hiperparámetros de forma razonable	1
1.2. Use el test de Wilcoxon de comparación de dos algoritmos sobre N problemas y aplíquelo a dos de los algoritmos anteriores. Obtenga el rango de Friedman para cada clasificador y configuración y represente gráficamente los resultados. Aplique el test de Iman-Davenport sobre los tres clasificadores	2
1.3. Compare el mejor método según el rango medio de Friedman con el resto de métodos usando el procedimiento de Holm . . .	4
1.4. Compare los métodos por parejas usando el procedimiento de Bonferroni-Dunn	4
1.5. Para uno de los clasificadores elegidos utilice una validación de los hiperparámetros con grid search y compare su rendimiento con el método con hiperparámetros fijados a priori	5
2. Preguntas práctica 2.2	6
2.1. Para cada uno de estos conjuntos aplique el método RIPPER del módulo “wittgenstein”. Evalúe los modelos obtenidos y compárelos con un árbol de decisión. Trate de interpretar los modelos obtenidos	6
2.2. Seleccione al menos 10 conjuntos de datos. Genere un clasificador SVM con un kernel lineal y un valor fijo de $C=1$. Compare el rendimiento con un árbol de decisión.	7
2.3. Utilice el método GridSearchCV para obtener los mejores hiperparámetros. Compare el rendimiento de los resultados usando validación cruzada de hiper-parámetros y los métodos usando valores fijos	8
3. Preguntas práctica 2.3	9
3.1. Aplique el método base a cada uno de los conjuntos y anote los resultados obtenidos.	9

3.2.	Aplique el método de combinación de clasificadores Bagging a cada uno de los conjuntos y anote los resultados obtenidos . . .	10
3.3.	Seleccione dos algoritmos de Boosting y aplique estos algoritmos a cada uno de los conjuntos y anote los resultados obtenidos	11
3.4.	Compare si hay diferencias significativas entre ellos usando el test de Iman-Davenport. Si es así, aplique el procedimiento de Wilcoxon para comparar cada método de agrupación con el clasificador base	12
3.5.	Enuncie las conclusiones del estudio indicando la influencia del clasificador base en el rendimiento de las agrupaciones de clasificadores	12
4.	Preguntas práctica 2.4	13
4.1.	Aplique el clasificador base a cada uno de los conjuntos y anote los resultados obtenidos	13
4.2.	Aplique los métodos multiclase one-vs.-one (OVO), one-vs.all (OVA) y error correcting output codes (ECOC) a cada uno de los conjuntos de datos y anote los resultados obtenidos.	14
4.3.	Compare si hay diferencias significativas entre ellos usando el test de Iman-Davenport. Si es así, aplique el procedimiento de Wilcoxon para comparar cada método multiclase con el clasificador base y los diferentes métodos entre ellos	16
4.4.	Conclusiones	16

Índice de figuras

1.	Precisión usando validación cruzada	2
2.	Rango de Friedman	3
3.	Ripper vs árbol de decisión	6
4.	Precisión del árbol de decisión y SVM	7
5.	Precisión clasificadores base	9
6.	Precisión clasificadores Bagging	10
7.	Precisión clasificadores Boosting	11
8.	Comparativa de métodos	12
9.	Precisión método base	13
10.	Precisión método OvO	14
11.	Precisión método OvR	15
12.	Precisión método ECOC	15

Índice de tablas

1.	Resultados aplicando validación cruzada	1
2.	Resultados del test de Wilcoxon	2
3.	Resultados del test de Friedman	3
4.	Resultados del test de Iman Davenport	4
5.	Comparativas de clasificadores	4
6.	Comparativas usando Bonferroni-Dunn	5
7.	Comparación entre tipos de parametros	5
8.	Comparación entre Ripper y árbol de decisión	6
9.	Precisión del árbol de decisión y SVM	7
10.	Mejores parametros y precisión SVM	8
11.	Clasificadores base	9
12.	Clasificadores Bagging	10
13.	Clasificadores Bagging	11
14.	Precisión con el clasificador base	13
15.	Resultados One vs One	14
16.	Resultados One vs One	16
17.	Resultados ECOC	16

En este documento se recogerán los resultados obtenidos realizando las pruebas indicadas en el guión de la práctica 2, así como el respectivo análisis de esos resultados. No se hará mención a las implementaciones en código necesarias para resolver cada problema.

1. Preguntas práctica 2.1

1.1. Seleccione como método para obtener el error la validación cruzada de 10 particiones o el método hold out. Seleccione dos métricas de las estudiadas en teoría. Entrene cada clasificador seleccionado y anote el valor de las métricas estudiadas. Fije los hiperparámetros de forma razonable

Se han usado como clasificadores arboles de decisión, kNN y maquinas de vectores soporte, usando en cada uno validación cruzada de 10 particiones. Como datasets se han usado *Iris*, *Diabetes*, *Ionosphere*, *Segment* y *Wine*. Como métricas, se usarán las mas sencillas posibles, la precisión del clasificador y el error cometido por este. Los resultados de ambas métricas, para cada clasificador están recogidas en la tabla 1. Además, se puede visualizar la diferencia de precisión de cada clasificador de forma gráfica en la figura 1.

Como hiperparámetros se han fijado *gini* como criterio para el árbol de decisión, 5 vecinos en el kNN y un kernel lineal con $C = 1$ para el SVM.

Dataset	Precisión Arbol	Error árbol	Precisión kNN	Error kNN	Precisión SVM	Error SVM
Iris	96 %	4 %	96.666 %	3.333 %	97.333 %	2.666 %
Diabetes	70.965 %	29.034 %	72.137 %	27.862 %	76.696 %	23.303 %
Ionosphere	88.888 %	11.111 %	82.341 %	17.658 %	86.904 %	13.095 %
Segment	96.066 %	3.933 %	93.066 %	6.933 %	96.4 %	3.599 %
Wine	88.169 %	11.830 %	67.549 %	32.450 %	95.555 %	4.444 %

Tabla 1: Resultados aplicando validación cruzada

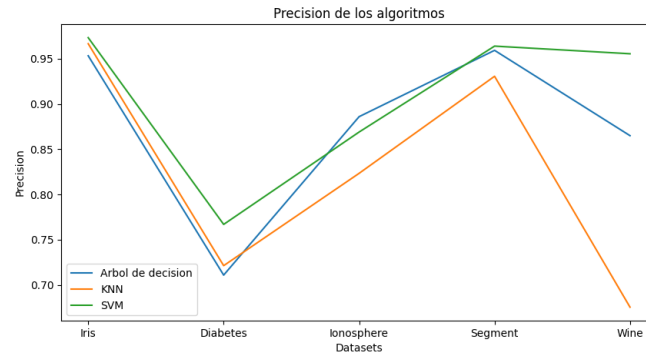


Figura 1: Precisión usando validación cruzada

1.2. Use el test de Wilcoxon de comparación de dos algoritmos sobre N problemas y aplíquelo a dos de los algoritmos anteriores. Obtenga el rango de Friedman para cada clasificador y configuración y represente gráficamente los resultados. Aplique el test de Iman-Davenport sobre los tres clasificadores

Para esta prueba, se usarán 4 datasets, *Iris*, *Wine*, *Breast Cancer* y *Digits*. Primeramente, el test de Wilcoxon comparando el árbol de decisión con el kNN proporciona los resultados presentes en la tabla 2. Podemos apreciar que no hay una diferencia significativa en ninguno de los casos, pues no llegamos a un valor de P menor a 0.05.

Dataset	Puntaje de Wilcoxon	Valor de P
Iris	0.0	0.3173
Wine	68.0	0.6547
Breast Cancer	18.0	0.1316
Digits	1336.5	0.1727

Tabla 2: Resultados del test de Wilcoxon

Usando los mismos datasets y como clasificadores el arbol de decisión, kNN y SVM, realizamos el test de Friedman. Los resultados se recogen en la tabla 3. Además, el rango de Friedman se puede ver de forma gráfica en la figura 2.

Por último, se aplica el test de Iman-Dravenport sobre los 3 clasificadores. Los resultados se recogen en la tabla 4.

Dataset	Rango de Friedman	Valor de P
Iris	<i>0.9999</i>	<i>0.6065</i>
Wine	<i>1.3454</i>	<i>0.5103</i>
Breast Cancer	<i>2.5333</i>	<i>0.2817</i>
Digits	<i>3.2048</i>	<i>0.2014</i>

Tabla 3: Resultados del test de Friedman

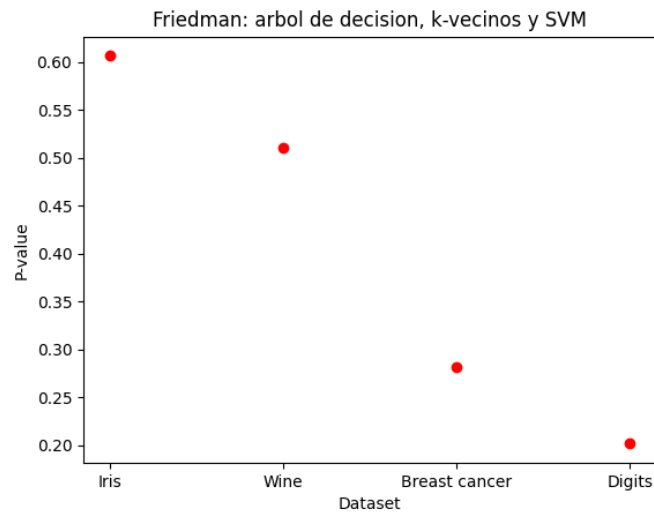


Figura 2: Rango de Friedman

Como breve conclusión, podemos apreciar que ninguno de los test muestra diferencias significativas entre clasificadores, pues en ninguno se obtiene un valor de $p < 0,05$.

Dataset	Puntaje de Iman Davenport	Valor de P
Iris	1028.0	0.8972
Wine	1494.5	0.8127
Breast Cancer	15048.0	0.5697
Digits	148917.5	0.2014

Tabla 4: Resultados del test de Iman Davenport

1.3. Compare el mejor método según el rango medio de Friedman con el resto de métodos usando el procedimiento de Holm

Para esto, podemos comparar los clasificadores con respecto a cada dataset. Generalmente, se aprecia que el mejor clasificador casi siempre es el kNN. Las comparativas se recogen en la tabla 5.

Dataset	Mejor clasificador	Valor de P
Iris	<i>kNN</i>	0.6065
Wine	<i>kNN</i>	0.6065
Breast Cancer	<i>kNN</i>	0.7903
Digits	<i>Árbol de decisión</i>	0.0149

Tabla 5: Comparativas de clasificadores

1.4. Compare los métodos por parejas usando el procedimiento de Bonferroni-Dunn

Se comparan todos los clasificadores en los distintos datasets, para a par. Los resultados se recogen en la tabla 6.

Como en las demás comparativas, no podemos apreciar ninguna diferencia significativa, pues ningún valor de $P < 0,05$.

Dataset	Clasificadores	Valor de P
Iris	<i>árbol vs kNN</i>	<i>0.8938</i>
Iris	<i>árbol vs SVM</i>	<i>1.0</i>
Iris	<i>kNN vs SVM</i>	<i>0.8938</i>
Wine	<i>árbol vs kNN</i>	<i>0.9186</i>
Wine	<i>árbol vs SVM</i>	<i>0.7338</i>
Wine	<i>kNN vs SVM</i>	<i>0.8093</i>
Breast Cancer	<i>árbol vs kNN</i>	<i>0.5693</i>
Breast Cancer	<i>árbol vs SVM</i>	<i>0.7316</i>
Breast Cancer	<i>kNN vs SVM</i>	<i>0.8209</i>
Digits	<i>árbol vs kNN</i>	<i>0.8620</i>
Digits	<i>árbol vs SVM</i>	<i>0.8112</i>
Digits	<i>kNN vs SVM</i>	<i>0.9465</i>

Tabla 6: Comparativas usando Bonferroni-Dunn

1.5. Para uno de los clasificadores elegidos utilice una validación de los hiperparámetros con grid search y compare su rendimiento con el método con hiperparámetros fijados a priori

Como clasificador se usará el árbol de decisión. Las comparativas entre la precisión obtenida usando los mejores hiperparámetros obtenidos por Grid-Search y la precisión obtenida usando hiperparametros fijados a priori se recogen en la tabla 7. La diferencia es obvia, al fijar los parametros a priori, corremos el riesgo de mal entrenar el modelo y obtener peores resultados. Como hiperparametros fijados a priori, se ha usado *gini* como criterio y sin limite de profundidad.

Dataset	Mejores Parametros	Precisión A priori	Precisión GridSearch
Iris	Criterio: entropy y profundidad = 10	<i>0.9555</i>	<i>0.9555</i>
Wine	Criterio: entropy y profundidad = 10	<i>0.9629</i>	<i>0.9444</i>
Breast Cancer	Criterio: entropy y profundidad = 10	<i>0.9415</i>	<i>0.9064</i>
Digits	Criterio: entropy y profundidad = 10	<i>0.8537</i>	<i>0.8722</i>

Tabla 7: Comparación entre tipos de parametros

2. Preguntas práctica 2.2

2.1. Para cada uno de estos conjuntos aplique el método RIPPER del módulo “wittgenstein”. Evalúe los modelos obtenidos y compárelos con un árbol de decisión. Trate de interpretar los modelos obtenidos

Para esta prueba, 3 datasets se han usado, *Iris*, *Breast Cancer* y *Titanic*. La comparativa de precisiones del árbol de decisión y Ripper se recogen en la tabla 8, y se visualiza de forma gráfica en la figura 3.

Dataset	Precisión Ripper	Precisión Árbol
Iris	94 %	98 %
Breast Cancer	99.468 %	91.4893 %
Titanic	81.0169 %	60 %

Tabla 8: Comparación entre Ripper y árbol de decisión

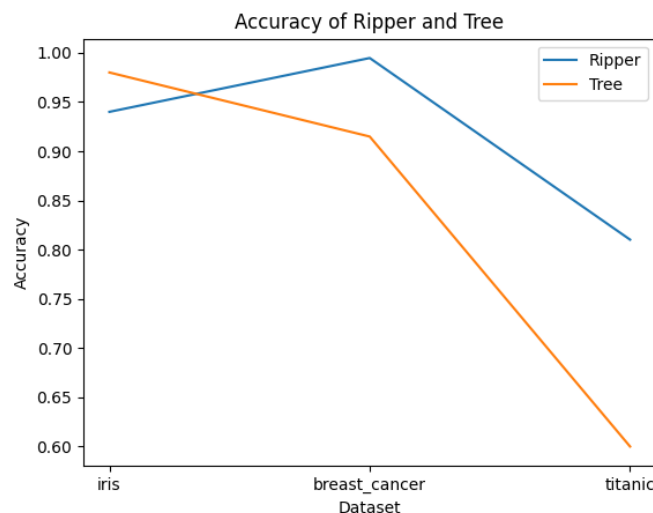


Figura 3: Ripper vs árbol de decisión

Podemos ver la diferencia entre modelos, viendo que para el dataset Iris, siendo sencillo, funciona mejor con el árbol de decisión, pero para dataset mas grandes, el método Ripper obtiene resultados muchos mejores.

2.2. Seleccione al menos 10 conjuntos de datos. Genere un clasificador SVM con un kernel lineal y un valor fijo de $C=1$. Compare el rendimiento con un árbol de decisión.

En este caso, se han seleccionado solamente 6 datasets, pues a la hora del análisis, se llegan a las mismas conclusiones, ahorrando tiempo. Se han usado los datasets *Iris*, *CPU*, *Diabetes*, *Ionosphere*, *Segment* y *Wine*. Los resultados de precisión, tanto del árbol de precisión como del SVM se recogen en la tabla 8. Se puede visualizar gráficamente en la figura 4.

Dataset	Precisión Árbol	Precisión SVM
Iris	94 %	98 %
CPU	0 %	7.9365 %
Diabetes	68.8311 %	76.6233 %
Ionosphere	87.7358 %	89.6226 %
Segment	95.7777 %	95.5555 %
Wine	100 %	99.4382 %

Tabla 9: Precisión del árbol de decisión y SVM

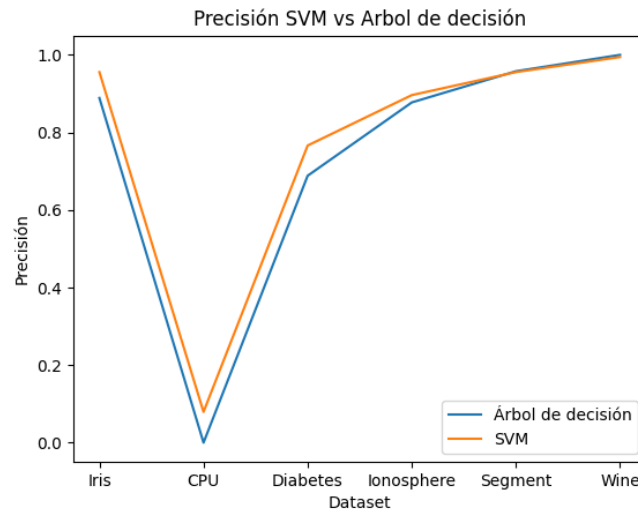


Figura 4: Precisión del árbol de decisión y SVM

2.3. Utilice el método GridSearchCV para obtener los mejores hiper-parámetros. Compare el rendimiento de los resultados usando validación cruzada de hiper-parámetros y los métodos usando valores fijos

Para este análisis, solo se va a comparar 3 datasets, *Iris*, *Ionosphere* y *Breast Cancer*. Como conclusión, obviamente los peores resultados se pueden llegar a obtener usando hiperparámetros seleccionados a priori, aunque exista la posibilidad de seleccionar la combinación óptima de hiperparámetros completamente a priori. Usando validación cruzada se puede mejorar los resultados al igual que usando el método GridSearchCV, basándose este método en la validación cruzada.

Dataset	Kernel Usado	Valor gamma	Valor C	Precisión SVM
Iris	<i>linear</i>	-	<i>1</i>	<i>94.2857 %</i>
Ionosphere	<i>rbf</i>	<i>10</i>	<i>0.1</i>	<i>94.6938 %</i>
Breast Cancer	<i>linear</i>	-	<i>10</i>	<i>94.9746 %</i>

Tabla 10: Mejores parametros y precisión SVM

3. Preguntas práctica 2.3

3.1. Aplique el método base a cada uno de los conjuntos y anote los resultados obtenidos.

Para esta prueba se usarán 6 datasets, *Iris*, *CPU*, *Ionosphere*, *Segment Challenge*, *Wine* y *Breast Cancer*. La comparativa entre ambos clasificadores se recoge en la tabla 12. Se puede visualizar de forma gráfica en la figura 5.

Dataset	Precisión árbol	Precisión SVM
Iris	95.3333 %	97.3333 %
Diabetes	71.0902 %	76.6968 %
Ionosphere	88.0555 %	86.9047 %
Segment Challenge	96.3333 %	95.5555 %
Wine	85.3595 %	95.5555 %
Breast Cancer	91.7418 %	95.4323 %

Tabla 11: Clasificadores base

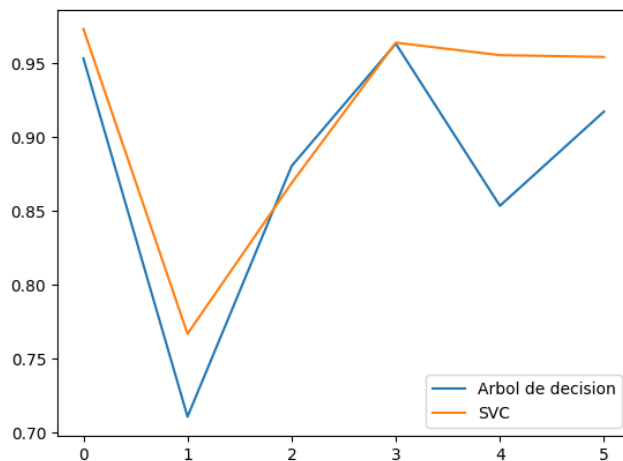


Figura 5: Precisión clasificadores base

3.2. Aplique el método de combinación de clasificadores Bagging a cada uno de los conjuntos y anote los resultados obtenidos

Se usarán los mismos datasets que en el apartado anterior, Los resultados se recogen en la tabla 13, y se puede visualizar en la figura 6.

Dataset	Precisión árbol	Precisión SVM
Iris	94 %	95.3333 %
Diabetes	73.1766 %	71.8626 %
Ionosphere	93.4285 %	87.7380 %
Segment Challenge	96.8 %	94.6666 %
Wine	97.2222 %	96.6339 %
Breast Cancer	94.7305 %	95.6077 %

Tabla 12: Clasificadores Bagging

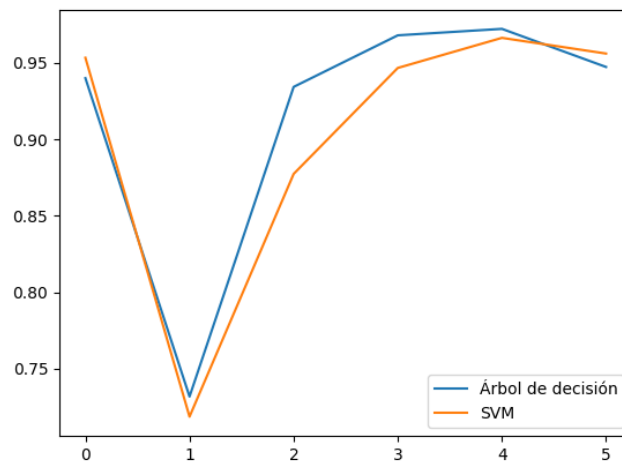


Figura 6: Precisión clasificadores Bagging

3.3. Seleccione dos algoritmos de Boosting y aplique estos algoritmos a cada uno de los conjuntos y anote los resultados obtenidos

Se usarán los mismos datasets, y como algoritmos de boosting, *AdaBoost* y *GradientBoost*. Los resultados se recogen en la tabla 13 y de visualizar en la figura 7.

Dataset	Precisión árbol	Precisión SVM
Iris	95.3333 %	96 %
Diabetes	75.524 %	75.9107 %
Ionosphere	92.0158 %	92.3015 %
Segment Challenge	97.9999 %	44.0666 %
Wine	98.3333 %	91.6013 %
Breast Cancer	96.3095 %	96.1403 %

Tabla 13: Clasificadores Bagging

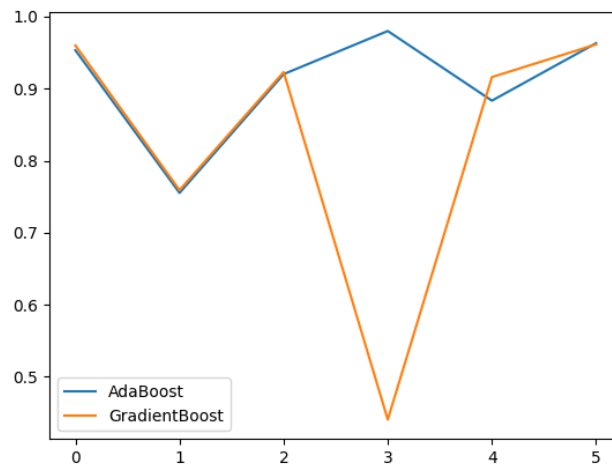


Figura 7: Precisión clasificadores Boosting

3.4. Compare si hay diferencias significativas entre ellos usando el test de Iman-Davenport. Si es así, aplique el procedimiento de Wilcoxon para comparar cada método de agrupación con el clasificador base

Tras aplicar el test de Iman-Davenport sobre todas las combinaciones de clasificadores base, usando Bagging y los dos métodos de Boosting, se llega a la conclusión de que para cada respectivo dataset no hay diferencias significativas en cuanto a clasificadores.

3.5. Enuncie las conclusiones del estudio indicando la influencia del clasificador base en el rendimiento de las agrupaciones de clasificadores

Como breve conclusión, es obvio que el clasificador base es lo que mas afecta a los resultados. Dentro de este clasificador base, se debe tener en cuenta que dataset es distinto, funcionando mejor o peor con ciertos clasificadores, asi como se debe tener en cuenta que la selección de hiperparámetros es también un factor determinante en los resultados. Por último, podemos ver gráficamente una comparativa entre todas los métodos en la figura 8.

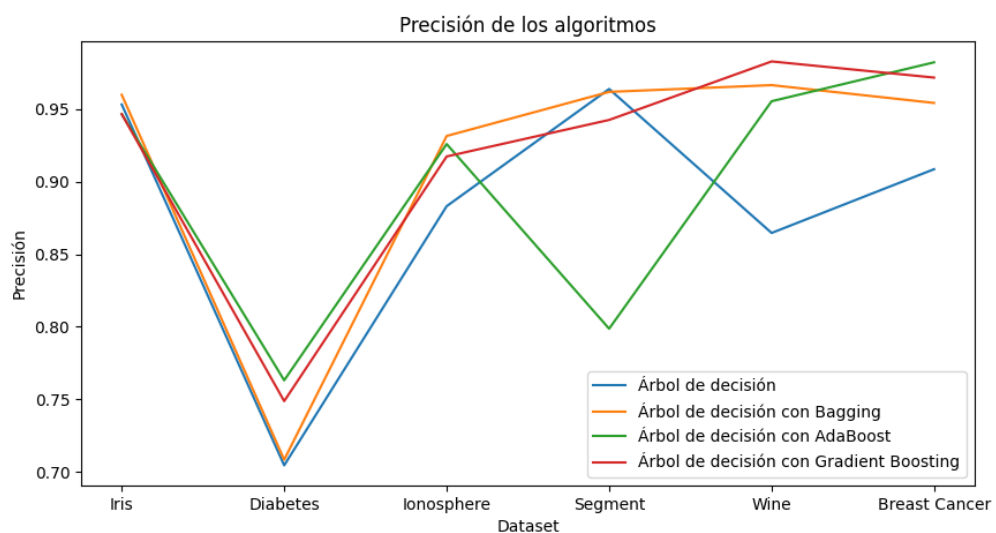


Figura 8: Comparativa de métodos

4. Preguntas práctica 2.4

4.1. Aplique el clasificador base a cada uno de los conjuntos y anote los resultados obtenidos

Como clasificar se usará un SVC (Support Vector Classifier) y como datasets, *Iris*, *Diabetes*, *Ionosphere*, *Segment*, *Wine* y *Breast Cancer*. Los resultados obtenidos con el clasificador base, usando un kernel lineal, se recogen en la tabla 14. Se puede apreciar la comparativa entre datasets en la figura 9.

Dataset	Precisión SVC
Iris	100 %
Diabetes	76.6233 %
Ionosphere	82.0754 %
Segment Challenge	96 %
Wine	99.438 %
Breast Cancer	96.6608 %

Tabla 14: Precisión con el clasificador base

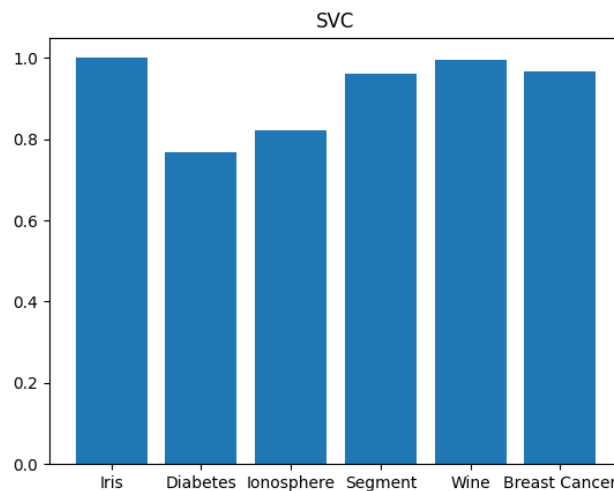


Figura 9: Precisión método base

4.2. Aplique los métodos multiclase one-vs.-one (OVO), one-vs.all (OVA) y error correcting output codes (ECOC) a cada uno de los conjuntos de datos y anote los resultados obtenidos.

Para el método One vs One, los resultados se recogen en la tabla 15 y se puede visualizar gráficamente en la figura 10.

Dataset	Precisión O v O
Iris	93.3333 %
Diabetes	77.0562 %
Ionosphere	83.0188 %
Segment Challenge	95.1111 %
Wine	99.4382 %
Breast Cancer	96.6608 %

Tabla 15: Resultados One vs One

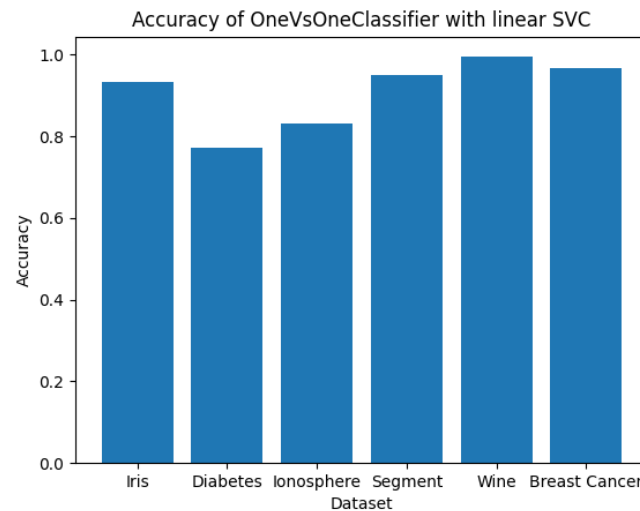


Figura 10: Precisión método OvO

Para el método One vs Rest, los resultados se recogen en la tabla 16 y se puede visualizar gráficamente en la figura 11.

Para el método ECOC, los resultados se recogen en la tabla 17 y se puede visualizar gráficamente en la figura 12.

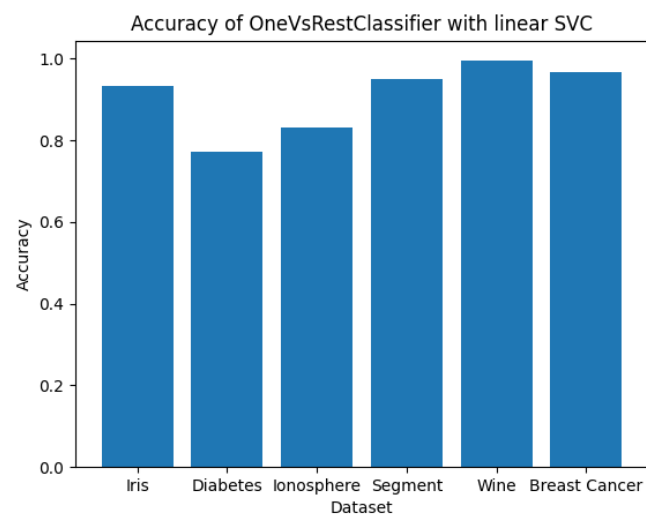


Figura 11: Precisión método OvR

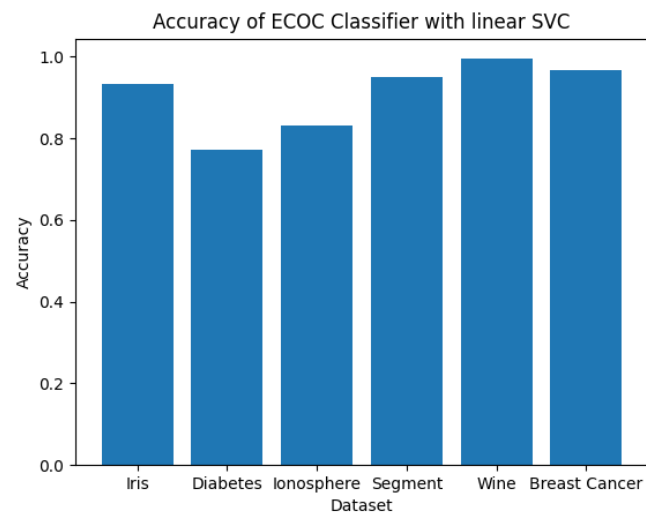


Figura 12: Precisión método ECOC

Dataset	Precisión O v R
Iris	<i>86.6666 %</i>
Diabetes	<i>77.0562 %</i>
Ionosphere	<i>83.0188 %</i>
Segment Challenge	<i>91.5555 %</i>
Wine	<i>99.4382 %</i>
Breast Cancer	<i>96.6608 %</i>

Tabla 16: Resultados One vs One

Dataset	Precisión ECOC
Iris	<i>80 %</i>
Diabetes	<i>77.0562 %</i>
Ionosphere	<i>83.0188 %</i>
Segment Challenge	<i>77.5555 %</i>
Wine	<i>99.4382 %</i>
Breast Cancer	<i>96.6608 %</i>

Tabla 17: Resultados ECOC

4.3. Compare si hay diferencias significativas entre ellos usando el test de Iman-Davenport. Si es así, aplique el procedimiento de Wilcoxon para comparar cada método multiclase con el clasificador base y los diferentes métodos entre ellos

Tras realizar el test de Iman-Davenport comparando todos los métodos por pares, y repitiendo las pruebas para cada datasets, estos tests afirman que no hay diferencias significativas entre métodos en ninguno de los casos, pues en ninguno se obtiene ningun valor de $P < 0,05$.

4.4. Conclusiones

Como conclusiones, podemos destacar que, como en los casos anteriores, el método influye, pero el dataset escogido y el clasificador base usado influye aún más. Dentro de los 3 métodos multiclase probados, los 3 arrojan resultados similares, y en ciertos casos el mismo, aunque se puede destacar como mejor método el One vs One, y el peor, el método ECOC.