

PRÁCTICA 5:

COMPARACIÓN DE VARIOS MÉTODOS DE APRENDIZAJE SOBRE VARIOS CONJUNTOS DE DATOS

**MINERÍA DE DATOS
PATRICIA AGUADO LABRADOR**

CONJUNTOS DE DATOS

Soybean	Ionosphere	Vote	Diabetes
683 instancias 36 atributos (35+c) 19 clases	351 instancias 35 atributos (34+c) 2 clases	435 instancias 17 atributos (16+c) 2 clases	768 instancias 9 atributos (8+c) 2 clases
Labor	Glass	Segment-test	Breast Cancer
57 instancias 17 atributos (1+c) 2 clases	214 Instancias 10 atributos (9+c) 6 clases	810 instancias 20 atributos (19+c) 7 clases	286 instancias 10 atributos (9+c) 2 clases
Credit-g	Iris	Car Evaluation	Balance-scale
1000 instancias 21 atributos (20+c) 7 clases	150 instancias 5 atributos (4+c) 3 clases	1728 instancias 7 atributos (6+c) 4 clases	625 instancias 5 atributos (4+c) 2 clases

Iris:

Conjunto de datos multivariante que contiene 50 muestras de cada una de las tres especies de Iris. Los datos recogen 4 rasgos de cada muestra: el largo y Ancho, del sépalo y pétalo, en centímetros. Atributos numéricos. Las clases son categóricas y corresponden a las tres especies.

<https://archive.ics.uci.edu/ml/datasets/Iris>

Car Evaluation:

Conjunto de datos que recoge las características de un coche en base a tres características que son precio, tecnología y confort. Los 6 atributos de cada Instancia son: precio, mantenimiento, puertas, personas, arranque de seguridad y seguridad. Atributos categóricos y numéricos. Las clases son de valor numérico y corresponden a la evaluación del automóvil.

<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

Balance-scale

Conjunto de datos generado para modelar resultados experimentales. Cada ejemplo se clasifica en función de la inclinación de una balanza (clase nominal). Los atributos son numéricos y hacen referencia a los pesos de ambos lados así como las distancias de ambos lados.

[https://github.com/renatopp/arff-datasets/blob/master/classification/balance.s
cale.arff](https://github.com/renatopp/arff-datasets/blob/master/classification/balance.scale.arff)

TEST DE SIGNOS

Tasas de error %		
	NB	J48
Soybean	7,04	8,49
Ionosphere	17,38	8,54
Vote	9,86	3,67
Diabetes	23,69	26,03
Labor	10	26,33
Glass	50,48	33,25
Segment-test	13,21	6,54
Breast Cancer	28,3	24,46
Credit-g	24,4	29,2
Iris	5,33	4
Car Evaluation	14,47	7,52
Balance-scale	8,64	36,8

Victorias:	5	7
------------	---	---

Tenemos 12 conjuntos de datos, para considerar que un algoritmo es significativamente mejor que otro necesitará ganar:

10 veces para $\alpha=0.05$

9 veces para $\alpha=0.1$

Como ninguno supera estas cifras, no podemos asegurar que ninguno de los dos algoritmos de clasificación sea mejor que el otro.

RANKINGS

	Tasas de error %					Rankings				
	SMO	3NN	NB	J48	OneR	SMO	3NN	NB	J48	OneR
Soybean	6,94	8,84	7,32	9,58	59,71	1	3	2	4	5
Ionosphere	11,91	13,67	17,67	10,94	18,8	2	3	4	1	5
Vote	4,14	6,62	9,89	3,59	4,37	2	4	5	1	3
Diabetes	22,99	26,28	24,58	27,39	27,81	1	3	2	4	5
Labor	10,88	8,3	8,64	20,24	25,58	3	1	2	4	5
Glass	42,89	32,05	50,58	33,07	44,95	3	1	5	2	4
Segment-test	7,78	7,85	13,56	6,12	39,06	2	3	4	1	5
Breast Cancer	30,7	27,14	26,85	26,85	32,8	4	3	1,5	1,5	5
Credit-g	24,5	28,04	24,9	27,98	33,5	1	4	2	3	5
Iris	3,6	5,2	4,93	5,33	6,93	1	3	2	4	5
Car Evaluation	6,54	6,52	14,53	8,48	29,98	2	1	4	3	5
Balance-scale	74,24	85,76	76,48	67,3	69,57	3	5	4	1	2
Ranking promedio:						2,08	2,83	3,125	2,45	4,5

TEST SOBRE RANKINGS

TEST DE IMAN Y DAVENPORT

$$X^2_F = \{ (12 \cdot N) / k \cdot (k + 1) \} \cdot \{ \sum R_j^2 - [k \cdot (k + 1)^2 / 4] \} =$$

$$\{ (12 \cdot 12) / 5 \cdot (5 + 1) \} \cdot \{ 48,35 - [5 \cdot (5 + 1)^2 / 4] \} =$$

$$\{ 4,8 \} \cdot \{ 3,35 \} = 16,08$$

$$F_F = \{ (N - 1) \cdot X^2_F \} / \{ N \cdot (k - 1) - X^2_F \} =$$

$$\{ (12 - 1) \cdot 16,08 \} / \{ 12 \cdot (5 - 1) - 16,08 \} =$$

$$\{ 176,88 \} / \{ 31,92 \} = 5,54$$

F con 4 y 44 grados de libertad = 2,584

Como $F_F > F$ se rechaza la hipótesis nula, es decir, los rankings promedios son significativamente distintos.

TEST POST-HOC

TEST DE NEMENYI

$q_\alpha = q_{0,05}$ para 5 clasificadores = 2,728

$CD = q_\alpha * \sqrt{[k*(k-1) / 6N]} =$

$2,728*(0,527) = 1,437$

	Ranking promedio	Intervalo	Diferente de
SMO	2,08	[0,643 , 3,517]	OneR
3NN	2,83	[1,393 , 4,237]	OneR
NB	3,125	[1,688 , 4,562]	-
J48	2,45	[1,013 , 3,887]	OneR
OneR	4,5	[3,063 , 5,937]	SMO, 3NN y J48

Dos métodos son diferentes si el ranking promedio del método con el que realizamos la comparación no pertenece al intervalo del método que se quiere comparar.

El único que es significativamente diferente al resto de métodos de clasificación de la tabla es OneR