

Multicomparación de Clasificadores

Sergio García Prado

November 19, 2016

Abstract

I. INTRODUCCIÓN

La comparación consistirá en dos partes principales: la primera de ellas se basa en un Test de Signos sobre 2 de los clasificadores para todos los conjuntos de datos, mientras que la segunda parte se refiere a la realización de un Ranking en el cuál participarán todos los clasificadores.

Para la realización de estas pruebas se ha utilizado Weka, que es una plataforma de software para el aprendizaje automático y la minería de datos escrita en Java, desarrollada en la Universidad de Waikato y distribuida como Software Libre.

Por lo tanto, lo primero es describir tanto los clasificadores como los conjuntos de datos que se utilizarán en los tests de clasificación:

I. Clasificadores

- **SVM con kernel lineal:**
- **3-NN:**
- **Naive Bayes:**
- **J48:**

II. Conjuntos de Datos

1. **Arrhythmia:**
2. **Diabetes:**
3. **Glass:**
4. **Ionosphere:**
5. **Iris:**
6. **Labor:**
7. **Seeds:**
8. **Segment Test:**
9. **Soybean:**
10. **Vote:**

II. TEST DE SIGNOS: SVM Y J48

El test de signos se ha realizado sobre los 10 conjuntos de datos aplicando validación cruzada de 10 particiones. Lo que se pretende con este test es contar el número de victorias de cada clasificador y otorgarle la victoria al que más número de veces haya sido seleccionado ganador. Los resultados obtenidos han sido los siguientes:

Datos	1	2	3	4	5	6	7	8	9	10
Ganador	SVM	SVM	J48	J48	SVM	SVM	SVM	J48	SVM	J48

Por lo tanto, los resultados de aplicar el test de signos son los siguientes:

$$Victorias(J48) = 4$$

$$Victorias(SVM) = 6$$

Entonces podemos concluir que para los conjuntos de datos utilizados, obtiene más victorias, por lo tanto en promedio genera mejores resultados, el clasificador **SVM**.

III. RANKING

Datasets	Resultados (Tasa Error)				Rankings			
	SVM (LK)	3-NN	Naive Bayes	J48	SVM (LK)	3-NN	Naive Bayes	J48
Arrhythmia	29.68%	41.72%	37.7%	33.85%	1	4	3	2
Diabetes	22.97%	26.28%	24.48%	27.37%	1	3	2	4
Glass	42.61%	32.05%	50.48%	33.54%	3	1	4	2
Ionosphere	11.91%	13.67%	17.67%	10.94%	2	3	4	1
Iris	3.6%	5.2%	4.67%	5.33%	1	3	2	4
Labor	10.88%	8.3%	5.88%	19.91%	3	2	1	4
Seeds	6.57%	7.14%	9.71%	9.14%	1	2	4	3
Segment Test	7.78%	7.85%	13.68%	6.1%	2	3	4	1
Soybean	6.94%	8.84%	7.32%	9.58%	1	3	2	4
Vote	4.14%	6.62%	9.89%	3.59%	2	3	4	1
Ranking Promedio					1.7	2.7	3	2.6

$$K = 4, N = 10, K - 1 = 3, N - 1 = 9, (K - 1)(N - 1) = 27 \alpha = 0.05 F_{3,27} = 2.960$$