

Multicomparación de Clasificadores

Sergio García Prado

November 20, 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

Realizaremos el test Friedman y el de Iman y Davenport:

Los valores que se utilizarán son los siguientes:

$$k = 4, N = 10, \alpha = 0.05, \chi^2_{0.05, (2)} = 5.991, F_{0.05, (3, 27)} = 2.960$$

Tenemos que calcular además el estadístico de Friedman:

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right] = \frac{120}{20} \left[(1.7^2 + 2.7^2 + 3^2 + 2.6^2) - \frac{100}{4} \right] = 6 * 0.94 = 5.64$$

El test de Friedman nos dice que no se rechaza la hipótesis nula por lo que no son significativamente diferentes. A pesar de ello, no es rechazada por muy poco, pero debido a la exigencia que pide este test ya sabemos que en el caso de Iman y Davenport tampoco será rechazada. A pesar de ello, realizaremos los cálculos:

Con este valor calculamos el estadístico necesario para el test de Iman y Davenport:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2} = \frac{9 * 5.64}{30 - 5.64} = 2.0837$$

Puesto que el valor de F_F no es mayor que el valor crítico de $F_{0.05,(3,27)}$ entonces no podemos rechazar la hipótesis nula, que consiste en asumir que los clasificadores son equivalentes.

Puesto que no son significativamente diferentes tanto con el test de Friedman (muy exigente) como con el de Iman y Davenport (menos exigente), no tiene sentido realizar el test post-hoc para conocer cuáles lo son ya que este test ya nos ha dicho que no lo es ninguno.