

Multicomparación de Clasificadores

Sergio García Prado

November 23, 2016

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, (SVM y J48) 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, 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:** Las máquinas de vector soporte son una técnica aprendizaje supervisado. Este modelo representa los puntos de muestra en el espacio, separando las clases en subespacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los puntos de las clases, mas cercanos al que se llama vector soporte. En este caso se ha utilizado un kernel lineal.
- **3-NN:** El método de los k vecinos más próximos es un método de clasificación supervisada no paramétrico. Estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de las muestras para calcular los k puntos medios. En este caso se ha utilizado 3 puntos.
- **Naive Bayes:** es un algoritmo de clasificación probabilístico basado en el teorema de Bayes y algunas hipótesis adicionales que facilitan la simplificación del problema. Estas hipótesis presuponen la independencia entre variables, de ahí es de donde proviene el apelativo *naive* (*ingenuo*) ya que esta presuposición no siempre es cierta. Es un clasificador que utiliza aprendizaje supervisado y utiliza el método de máxima verosimilitud.
- **J48:** es un algoritmo cuya función es generar un árbol de decisión. Su nombre original es *C4.5* pero la implementación de Weka se denomina *J48*. Para generar el árbol de decisión el árbol basa la selección de atributos en cada nodo según la entropía de los mismos con respecto a la clases en la que se desea clasificar las muestras. Es por tanto un algoritmo de aprendizaje supervisado.

II. Conjuntos de Datos

1. **Arrhythmia:** 452 muestras con 279 atributos y 16 clases.
2. **Diabetes:** 768 muestras con 8 atributos y 2 clases.
3. **Glass:** 214 muestras con 9 atributos y 7 clases.

4. **Ionosphere**: 351 muestras con 34 atributos y 2 clases.
5. **Iris**: 150 muestras con 4 atributos y 3 clases.
6. **Labor**: 57 muestras con 16 atributos y 2 clases.
7. **Seeds**: 210 muestras con 7 atributos y 3 clases.
8. **Segment Test**: 210 muestras con 19 atributos y 7 clases.
9. **Soybean**: 683 muestras con 35 atributos y 19 clases.
10. **Vote**: 435 muestras con 16 atributos y 2 clases.

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 de los tests se han obtenido a partir de la tasa de error estimada mediante validación cruzada de 10 particiones.

Los resultados obtenidos han sido los siguientes (siendo el conjunto de datos i el que se encuentra en dicha posición en el apartado anterior.):

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**. Las victorias que obtiene este clasificador son en los conjuntos de datos cuyos atributos son de carácter numérico, mientras que J48 destaca cuando son atributos categóricos.

III. RANKING

El ranking que se muestra en la tabla inferior muestra los resultados de obtener las tasas de error de cada clasificador en cada conjunto de datos. Para realizar el ranking promedio se ha realizado un ranking para conjunto de datos y una vez realizado esto, se han promediado los resultados. La tasa de error se ha obtenido mediante validación cruzada repetida de 5 particiones con 5 repeticiones.

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

El siguiente paso es comprobar si estos resultados son significativos según los tests de Friedman y de Iman y Davenport:

Los valores que se utilizarán en las fórmulas son los siguientes:

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

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.

Debido a 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.

En el caso del ranking promedio en el cual aparecen todos los clasificadores, el resultado es equivalente al test de signos para la máquina de vectores soporte y J48. El comportamiento de los otros dos algoritmos (Ibk y Naive Bayes) obtiene peores resultados en promedio. Las posiciones en el ranking no están muy diferenciadas entre si, es decir, para el número de datos y clasificadores, es decir, los resultados no son significativos (con otros conjuntos de datos podrían haberse posicionado de manera diferente en el ranking promedio). Como consecuencia de esto no podemos asegurar que ninguno obtenga mejores resultados que otros.