

Práctica metodología experimental: creación y evaluación de hipótesis

Alejandro Pulido Sánchez

1. Descripción de los conjuntos de datos

- **Soybean:** formado por 683 instancias formadas por 35 atributos, todos ellos de carácter nominal. La clase de destino puede tomar 19 valores distintos. El conjunto de datos se corresponde con instancias referidas a atributos de plantas y la clase de destino representa el tipo de planta.
- **Vote:** formado por 435 instancias formadas por 16 atributos, todos ellos de carácter nominal. La clase de destino puede tomar 2 valores distintos. El conjunto de datos se refiere a resultados de encuestas a ciudadanos estadounidenses para tratar de predecir si votarán al partido demócrata o republicano.

2. Descripción de los algoritmos (no hay que describir C4.5, es un estándar; no hay que describir Weka, pero sí indicar las herramientas que utilizáis)

- **50T:** Para obtener 50 instancias, utilizo la opción "Percentage split" de Weka. Para el conjunto de datos Soybean, utilizo el 7.325% y para el conjunto de datos Vote, utilizo el 11.5%
- **Hold Out:** En el método "Hold Out", divido el conjunto de datos en dos partes: una para entrenamiento y otra para prueba. Siguiendo la indicación del enunciado de utilizar $\frac{1}{3}$ y $\frac{2}{3}$, selecciono un 66%.
- **Hold Out Repetido:** En el método "Hold Out Repetido", genero más valores cambiando la semilla de selección de conjuntos en Weka por 2, 3 y 4, y calculo las tasas medias de error y otros parámetros.
- **Validación Cruzada:** En el método "Validación Cruzada", utilizo la opción "Cross-validation" de Weka y selecciono un parámetro de 10 particiones. En cada subconjunto de particiones, uso una parte como conjunto de prueba y el resto como conjunto de entrenamiento.
- **Validación Cruzada Repetida:** Para el método "Validación Cruzada Repetida", repito el proceso del "Hold Out Repetido", cambiando las semillas de selección de subconjuntos por 2, 3 y 4

3. Experimentos con las muestras de 50 instancias de cada conjunto de datos (ejercicio previo)

Datos	Algoritmo	Método: 50T, resto			
		Tasa error	Desviación estándar	Intervalo superior	Intervalo inferior
Soybean_50	J48	0,5024	0,0199	0,5054	0,4994
	Sin podar	0,5024	0,0199	0,5054	0,4994
Vote_50	J48	0,0468	0,0108	0,0489	0,0447
	Sin podar	0,0468	0,0108	0,0489	0,0447

4. Experimentos de hold out sin repetición

Datos	Algoritmo	Método: Hold out			
		Tasa error	Desviación estándar	Intervalo superior	Intervalo inferior
Soybean_50	J48	0,0949	0,0192	0,0998	0,0192
	Sin podar	0,1336	0,0223	0,1392	0,0223
Vote_50	J48	0,0271	0,0133	0,0313	0,0133
	Sin podar	0,0271	0,0133	0,0313	0,0133

5. Experimentos de hold out con repetición

Datos	Algoritmo	Tasa de error		
		2	3	4
Soybean_50	J48	0,1121	0,1078	0,1379
	Sin podar	0,1164	0,1293	0,1422
Vote_50	J48	0,0811	0,0541	0,0608
	Sin podar	0,0676	0,0541	0,0608

Datos	Algoritmo	Método: Hold out repetido			
		Tasa error	Desviación estándar	Intervalo superior	Intervalo inferior
Soybean_50	J48	0,1132	0,0208	0,1184	0,1079
	Sin podar	0,1304	0,0221	0,1360	0,1248
Vote_50	J48	0,0558	0,0189	0,0618	0,0498
	Sin podar	0,0524	0,0183	0,0582	0,0466

6. Experimentos de validación cruzada sin repetición

Datos	Algoritmo	Método: 10 XV			
		Tasa error	Desviación estándar	Intervalo superior	Intervalo inferior
Soybean_50	J48	0,0849	0,0082	0,0864	0,0834
	Sin podar	0,0838	0,0092	0,0855	0,0821
Vote_50	J48	0,0368	0,0114	0,0389	0,0347
	Sin podar	0,0368	0,0115	0,0389	0,0347

7. Experimentos de validación cruzada con repetición

Datos	Algoritmo	Tasa de error		
		2	3	4
Soybean_50	J48	0,0981	0,0908	0,0791
	Sin podar	0,1054	0,1025	0,0893
Vote_50	J48	0,0322	0,0368	0,0345
	Sin podar	0,0437	0,0414	0,0391

Datos	Algoritmo	Método: 10 XV			
		Tasa error	Desviación estándar	Intervalo superior	Intervalo inferior
Soybean_50	J48	0,0882	0,0053	0,0885	0,0880
	Sin podar	0,0953	0,0053	0,0955	0,0950
Vote_50	J48	0,0351	0,0043	0,0353	0,0349
	Sin podar	0,0403	0,0050	0,0405	0,0400

8. Tablas comparativas y discusión de resultados

- Soybean:

Algoritmo	50 instan. entrenam.	Hold out	Hold out repetido	10-XV	4 x 10 XV
J48					
Error	0,5024	0,0949	0,1132	0,0849	0,0882
Desviación	0,0199	0,0192	0,0208	0,0082	0,0053
Intervalo superior	0,5054	0,0998	0,1184	0,0864	0,0885
Intervalo inferior	0,4994	0,0192	0,1079	0,0834	0,0880
Sin podar					
Error	0,5024	0,1336	0,1304	0,0838	0,0953
Desviación	0,0199	0,0223	0,0221	0,0092	0,0053
Intervalo superior	0,5054	0,1392	0,1360	0,0855	0,0955
Intervalo inferior	0,4994	0,0223	0,1248	0,0821	0,0950

- Vote:

Algoritmo	50 instan. entrenam.	Hold out	Hold out repetido	10-XV	4 x 10 XV
J48					
Error	0,5024	0,0271	0,0558	0,0368	0,0351
Desviación	0,0199	0,0133	0,0189	0,0114	0,0043
Intervalo superior	0,5054	0,0313	0,0618	0,0389	0,0353
Intervalo inferior	0,4994	0,0133	0,0498	0,0347	0,0349
Sin podar					
Error	0,0468	0,0271	0,0524	0,0368	0,0403
Desviación	0,0108	0,0133	0,0183	0,0115	0,0050
Intervalo superior	0,0489	0,0313	0,0582	0,0389	0,0405
Intervalo inferior	0,0447	0,0133	0,0466	0,0347	0,0400

Según los resultados experimentales, podemos observar lo siguiente:

- La tasa de error varía significativamente según la regla utilizada para entrenar el modelo. La regla Hold Out repetida obtiene la tasa de error más alta (0.1132), mientras que la validación cruzada repetida obtiene la tasa de error más baja (0.0849).
- La desviación también varía entre las reglas. La validación cruzada repetida obtiene la desviación más baja (0.0082), mientras que la regla sin podar obtiene la desviación más alta (0.0223).
- Los intervalos superiores e inferiores también varían según la regla utilizada. La regla Hold Out repetida obtiene el intervalo superior más alto (0.1184) y la regla sin podar obtiene el intervalo superior más bajo (0.1360). La validación cruzada repetida obtiene el intervalo inferior más bajo (0.0834) y la regla Hold Out repetida obtiene el intervalo inferior más alto (0.1079).

En general, la validación cruzada repetida parece ser la mejor regla en términos de tasa de error y desviación, mientras que la regla sin podar es la peor. Sin embargo, los resultados pueden variar según las características de los datos y la tarea de clasificación específica. Por lo tanto, es importante realizar más evaluaciones y comparaciones para tomar una decisión informada sobre qué regla utilizar en una tarea de clasificación específica.

9. Preguntas sobre validación cruzada

¿Qué tasa de error se obtendría con el método 2?

Debería ser la misma.

¿Cómo espera que varíe la estimación de la varianza con el método 2 frente al método 1?

Se espera que su valor sea inferior.

¿Y los intervalos de confianza?

Se espera que sean más pequeños.

10. Referencias

Apuntes de la asignatura TAA.