

Errores de Aprendizaje Automático según el tipo de experimento

Juan Francisco Mier Montoto, EPI Gijón 22-23

—

Metodología

Tipos de experimentos

En este análisis se van a utilizar los siguientes tipos de experimentos:

- **Validación cruzada:** el conjunto total de datos se fracciona, normalmente en 10 partes, donde una de las partes se utiliza como conjunto de prueba y el resto como conjunto de entrenamiento. El entrenamiento se repite tantas veces como partes haya, de forma que cada subconjunto se utiliza nueve veces en el conjunto de entrenamiento y una en el conjunto de prueba.
- **Reescritura:** el conjunto completo de datos se utiliza tanto para entrenamiento como para pruebas.

Algoritmos

Para la práctica, se utilizan los siguientes algoritmos de clasificación:

- *OneR*
- *C4.5 / J48*
- *SMO (SVM)*

Resultados

A partir de los experimentos indicados y según la información aportada en el enunciado, se obtiene la siguiente tabla de errores (porcentajes de incorrectos):

	VALIDACIÓN CRUZADA			REESCRITURA		
Dataset	OneR	J48 (C4.5)	SMO (SVM)	OneR	J48 (C4.5)	SMO (SVM)
diabetes	28.48	25.51	23.20	23.57	15.89	22.53
iris	7.47	5.27	3.73	4.00	2.00	3.33
mushroom	1.48	0.00	0.00	1.48	0.00	0.00
zoo	57.41	7.39	3.95	0.00	0.99	0.00

Análisis

A la hora de analizar los resultados, llama la atención la gran diferencia que existe en el conjunto de *zoo* utilizando el algoritmo *OneR* entre **validación cruzada** y **reescritura**.

Estas diferencias entre ambos tipos se deben al tipo de datos de cada conjunto y el funcionamiento de los tipos de experimentos:

- En **zoo** se espera obtener el tipo de animal en base a diferentes atributos, como por ejemplo si es acuático o no, si pone huevos o no, etc. Utilizando **reescritura** se obtienen valores de error muy bajos, puesto que el sistema ha sido entrenado sobre todo el conjunto de datos y conoce la clasificación correcta. Por otra parte, **validación cruzada** entrena al sistema repetidas veces con conjuntos parciales, por lo que es más difícil que se clasifique correctamente el tipo de animal.
 - Siendo más concretos, la mayor diferencia se presenta al utilizar el algoritmo *OneR*, que escoge un solo atributo como medida para clasificar cada animal. De esta manera, se consigue muy buena precisión en reescritura pero terrible en validación cruzada.
 - Sin embargo, esto no significa que, una vez entrenado, el sistema responda con esas tasas de errores sobre otros conjuntos: predeciblemente, validación cruzada está mejor preparado para ello, mientras que reescritura no será capaz de clasificar correctamente.
- En otros conjuntos de datos, las diferencias entre tipos de experimentos son menores. En **mushroom**, solo se quieren clasificar las setas venenosas de las comestibles. Al reducir el número de posibles clasificaciones, las características de ambos tipos de experimentos no tienen tanta relevancia.
 - En *OneR*, al obtener la mayoría de ocurrencias de un solo atributo se clasifica de la misma manera pese a que en validación cruzada se reduzcan los conjuntos de entrenamiento. Por ello, se obtiene el mismo resultado en ambos tipos de experimento.

Queda en evidencia que *OneR* no es un algoritmo fiable de clasificación, ya que escoger un solo atributo no permite la correcta diferenciación de datos cuando no se utiliza el conjunto completo para el entrenamiento. Los otros algoritmos de clasificación son mucho más fiables y no presentan diferencias de error tan grandes entre tipos de experimentos, puesto que estudian los datos utilizando múltiples atributos.