Trabajo de Fin de Grado

Evaluación del estado del arte en técnicas estadísticas para el análisis comparativo de algoritmos de aprendizaje automático

Jacinto Carrasco Castillo

19 de septiembre de 2016

Planteamiento

Descripción del problema

La comparación de algoritmos debe realizarse de manera rigurosa para reducir la aleatoriedad asociada a los experimentos. Para realizar una correcta comparación es necesario:

- Selección de medida del rendimiento.
- Validación cruzada y remuestreo.
- Test estadísticos.

Objetivos

- Estudio de los test estadísticos disponibles.
- Integración de herramientas informáticas existentes para la aplicación de estos test.
- Comparación de las propiedades de los test.

Contenido matemático

Test paramétricos

Suponen que la muestra pertenece a una distribución conocida.

Test binomial para una muestra Comprobación de que el rendimiento para un problema sea igual a un valor θ_0.

***t*-test para muestras apareadas** Comparación de la media de dos muestras. Comparación de dos algoritmos para un conjunto de datos.

ANOVA Comparación de la media de múltiples algoritmos en múltiples problemas. Trata la varianza en un grupo, entre grupos y la combinación de ellas.

Test no paramétricos

Comparación con test paramétricos

- Los test no paramétricos no suponen la pertenencia de la distribución a ninguna familia de distribuciones.
- Las hipótesis para aplicar estos test son más generales.
- Si se dan las hipótesis necesarias para los test paramétricos, tienen una menor potencia.
- Si se disponen de pocos datos las hipótesis de los test paramétricos no suelen darse y los test no paramétricos suplen la falta de potencia con mayor precisión.

Test no paramétricos

Tipos de test no paramétricos:

- Test de aleatoriedad: basados en el número de rachas.
- Test de bondad del ajuste: Test chi cuadrado, Kolmogorov-Smirnov.
- Análisis del conteo de datos: Test de McNemar.
- Test basado en una muestra y muestras apareadas: Test de signo, test de rangos con signos de Wilcoxon.
- Análisis bidimensional de la varianza: Test de Friedman, modificación de Iman-Davenport, test de rangos alineados de Friedman, test de Quade.

Test no paramétricos

Procedimientos *post-hoc*

Los test que comparan múltiples algoritmos indican la existencia de diferencias entre ellos. Necesitan un test adicional para indicar dónde están estas diferencias.

- *p*-valores ajustados: Al realizar múltiples comparaciones aumenta la probabilidad de cometer un error de tipo I. Hay distintos métodos para ajustar los *p*-valores obtenidos.
  - Procedimientos de un paso: Bonferroni-Dunn.
  - Procedimiento descendente: Holm, Holland, Finner.
  - Procedimiento ascendente: Hochberg, Hommel, Rom.
  - Procedimiento en dos pasos: Li.

Test no paramétricos

Test basados en permutaciones

Son test no paramétricos. La hipótesis nula es que una muestra **x** proviene de una misma población.
Suponiendo H_0 cierta, los individuos de cada subconjunto de la muestra se puede intercambiar por los de otro subconjunto.

Definición (Principio de los test basados en permutaciones)

Si dos experimentos toman valores en el mismo espacio muestral Ω con distribuciones P_1, P_2 dan el mismo conjunto de datos, las inferencias condicionales a los datos usando el mismo estadístico deben ser la misma.

Test bayesianos

Comparación con THN

La inferencia bayesiana ajusta un modelo de probabilidad a los datos y obtiene una distribución sobre los parámetros del modelo. Las diferencias con los test de hipótesis nula son:

- Se evitan decisiones dicotómicas marcadas por α.
- Los THN no estiman la probabilidad de la hipótesis.
- Con suficientes datos se rechaza casi toda hip. nula.
- No se tiene en cuenta magnitud de la diferencia ni incertidumbre.
- No se obtiene información si no se rechaza la hipótesis nula.

Test bayesianos

t-test bayesiano correlado Comparación de dos dos algoritmos en un único conjunto de datos. Tiene en cuenta la correlación de los conjuntos de entrenamientos en CV. Se obtiene una distribución T de Student sobre la diferencia de las medias.

Test bayesiano de signo Versión bayesiana del test de signo. Se obtiene distribución sobre la probabilidad de que la diferencia entre algoritmos esté en la *rope*, a la izquierda o a la derecha.

Test bayesiano de rangos alineados Versión bayesiana del test de rangos alineados. No se obtiene una distribución clara de los parámetros, pero se puede obtener una muestra.

Contenido informático

rNPBST

Características del paquete

Característica	
Lenguaje	`R`
Test no paramétricos	Integrados a partir del paquete `JavaNPST` usando `rJava`.
Test bayesianos	Implementados `R`. Mayor eficiencia y reusabilidad que los paquetes disponibles. Incluye métodos para la representación gráfica de los resultados de estos test.
Datos de prueba	Se incluyen los resultados de 5 algoritmos sobre 29 conjuntos de datos para ejemplificar el uso del paquete.

rNPBST

Instalación

Paquete de `R` disponible en

`https://github.com/JacintoCC/TFG/tree/master/rNPBST`

Para la instalación, ejecutar donde se encuentre la carpeta con el software:

```
> R CMD build rNPBST
> R CMD INSTALL rNPBST
```

Aplicación de test

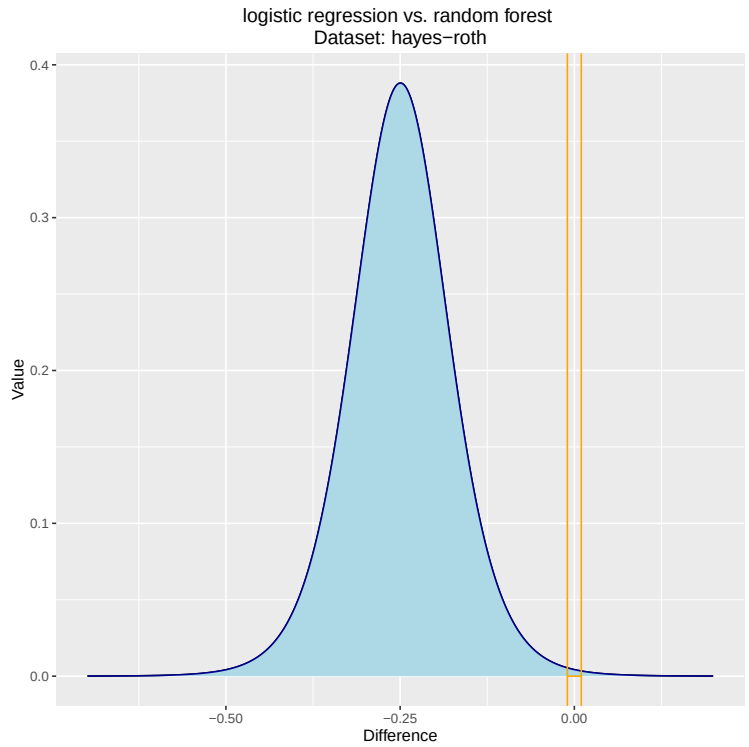
```
> data("results")
> ft <- friedman.test(results)
> ft$htest
    Friedman test
data: results
s = 2812.000, q = 39.056, p-value = 6.789e-08
```

Con estos resultados rechazaríamos la hipótesis nula de la equivalencia de los algoritmos. El test nos devuelve en el parámetro `report` también la suma del orden medio de cada algoritmo.

Aplicación de test

Test bayesianos - *t*-Test bayesiano correlado

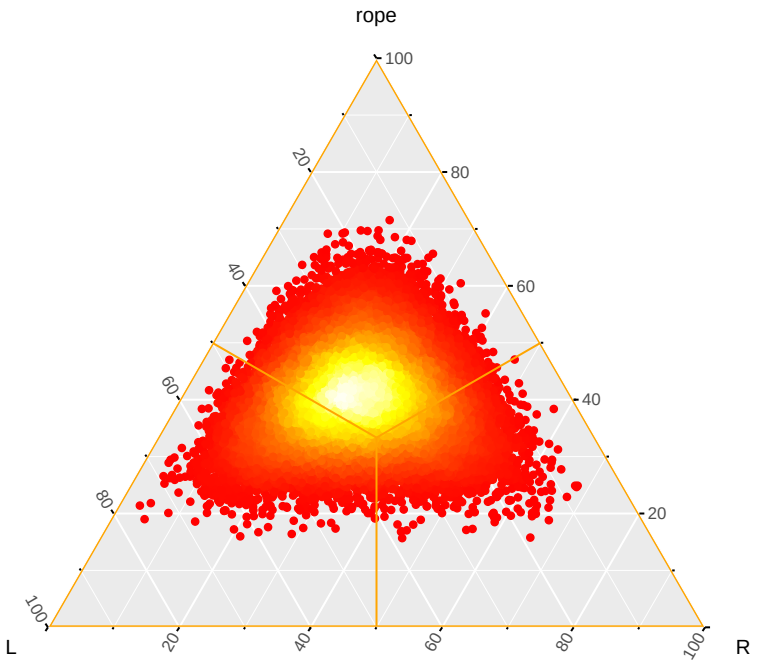
```
> dataset.index <- 13
> correlatedBayesianT.test(results.lr[dataset.index, ],
                           results.rf[dataset.index, ])
```



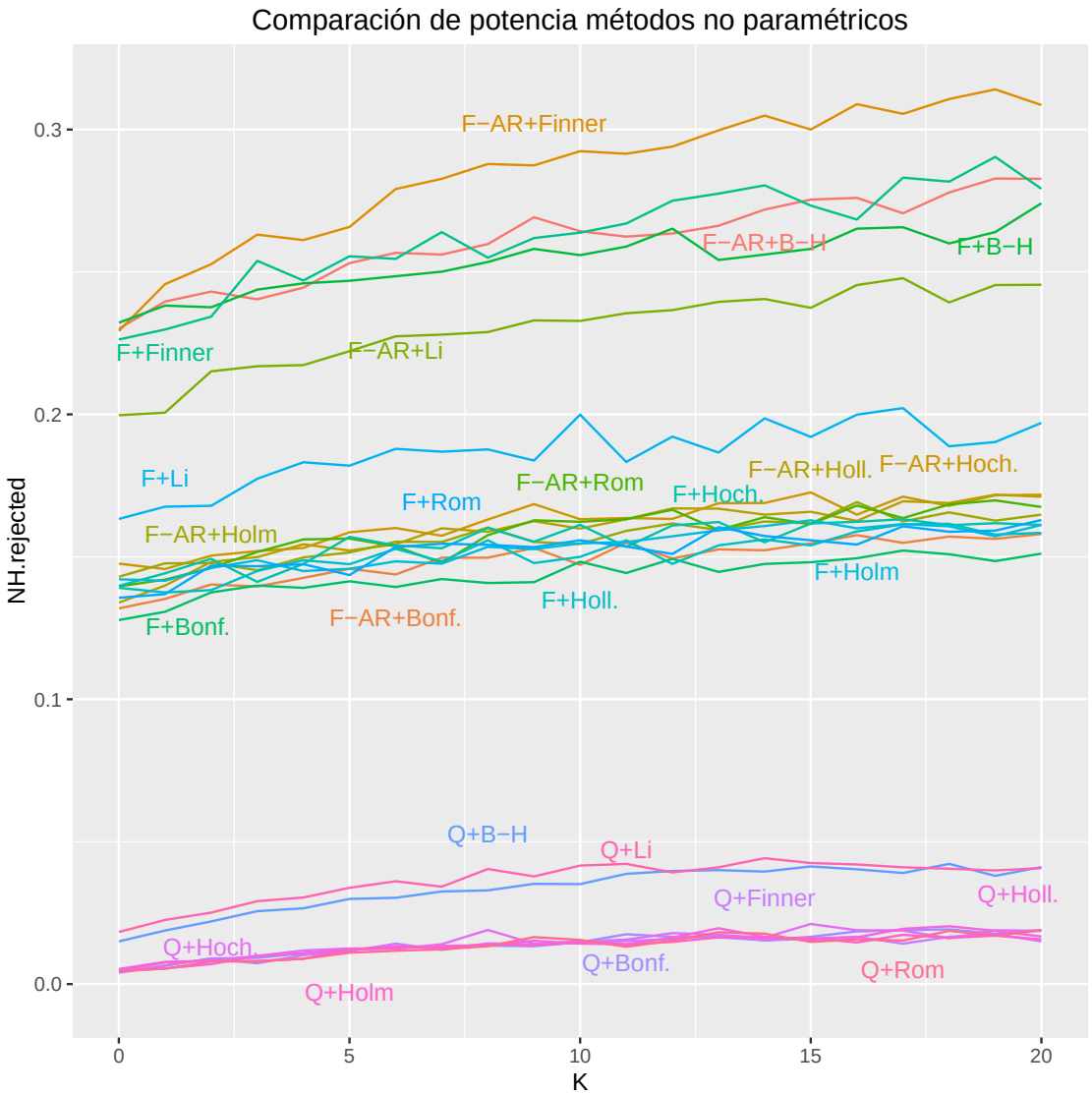
Aplicación de test

Test bayesianos - Test bayesiano de rangos con signo

```
> bayesianSignedRank.test(results$KNN,
                          results$neural.network)
```



Comparación de la potencia de test estadísticos



Conclusiones

Conclusiones

Se presentan las conclusiones obtenidas tras la realización del trabajo:

- Necesidad de comprobar las condiciones necesarias para la aplicación de test.
- Relevancia de los test basados en *rankings* para la comparación de múltiples algoritmos.
- Diferencias entre THN y test bayesianos.
- Importancia de disponer una única herramienta para la realización de test.

Trabajos futuros

- Seguir profundizando en nuevos métodos para realizar la comparación propuestos
- Complementar la biblioteca de `R` con estos nuevos métodos propuestos, test basados en permutaciones.
- Realizar un estudio más detallado sobre las propiedades de los test y métodos, incluir comparación de test bayesianos.

Principales fuentes bibliográficas

- *Evaluating Learning Algorithms: A Classification Perspective*, N. Japkowicz
- *Nonparametric Statistical Inference*, J.D. Gibbons y S. Chakraborti
- *Statistical Comparisons of Classifiers over Multiple Data Sets*, J. Demšar
- *An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons*, S. García y F. Herrera
- *Permutation tests for complex data: theory, applications and software*, F. Pesarin y L. Salmaso
- *Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis*, A. Benavoli *et. al*
- Documentación de los paquetes utilizados para la herramienta software.