



Ejercicio práctico nº 10

Enunciado sala 6

Pregunta 1

Dos artículos reportan el porcentaje de acierto alcanzado por dos algoritmos de clasificación, específicamente el Bayes ingenuo (C4) y el Bayes ingenuo oculto (C2), en diferentes conjuntos de prueba disponibles en el UCI Machine Learning Repository. ¿Es uno de los algoritmo mejor que el otro?

```
texto <- ("
Dataset          C2          Dataset          C4
'ecoli'           79.48      'page-blocks'    92.95
'primary-tumor'   47.52      'squash-unstored' 61.11
'pasture-production' 85.27      'mushroom'       95.27
'contact-lenses'  67.77      'segment'        90.74
'nursery'         93.72      'cmc'            50.49
'white-clover'    78.73      'soybean'        91.52
'monks1'          99.44      'credit'         85.67
'anneal'          97.44      'monks'          61.68
'solar-flare-C'   87.68      'postoperatie'   66.11
'kr-s-kp'         91.90      'grub-damage'    47.23
'tae'             43.82      'waveform'       79.30
'squash-stored'   57.44      --              --
")
datos <- read.table(textConnection(texto), header = TRUE, na.strings = "--")
```

Pregunta 2

Proponga un ejemplo novedoso (no mencionado en clase ni que aparezca en las lecturas dadas) en donde un estudio o experimento, relacionado con el alza que han experimentado las tasas de interés de los créditos en Chile, necesite utilizar una prueba de los rangos con signo de Wilcoxon, debido a problemas con la escala de la variable dependiente en estudio. Indique cuáles serían las variables/niveles involucrados en su ejemplo y las hipótesis nula y alternativa a contrastar.



Pregunta 3

El siguiente texto muestra porcentaje de acierto alcanzado por tres algoritmos de clasificación en diferentes conjuntos de prueba disponibles en el UCI Machine Learning Repository. Los algoritmos corresponden a C3: averaged one-dependence estimator (AODE), C6: locally weighted naive-Bayes y C7: random forest. ¿Existe un algoritmo mejor o peor que los otros?

```
texto <- ("  
Dataset      C3      C6      C7  
'pima-diabetes' 74.45  74.19  72.11  
'pendigits'   97.26  94.25  95.11  
'credit'       84.51  84.66  82.77  
'eucalyptus'   58.15  58.96  58.84  
'primary-tumor' 46.93  48.99  37.75  
'waveform'     84.36  83.06  79.12  
'solar-flare-X' 97.28  93.85  95.43  
'glass'        73.27  75.13  72.77  
'solar-flare-m' 87.36  86.43  84.90  
'hepatitis'    83.23  81.94  80.69  
'sonar'        80.70  80.23  77.80  
'page-blocks'  96.39  93.59  96.41  
'solar-flare-C' 87.98  87.36  85.49  
'yeast'        57.18  56.92  55.70  
'optdigits'    96.34  93.64  91.24  
'iris'         92.11  91.44  92.77  
")  
datos <- read.table(textConnection(texto), header = TRUE)
```

Pregunta 4

Proponga un ejemplo novedoso (no mencionado en clase ni que aparezca en las lecturas dadas) en donde un estudio o experimento, relacionado con el alza que han experimentado las tasas de interés de los créditos en Chile, necesite utilizar una prueba de Kruskal-Wallis, debido a problemas con la normalidad de los datos. Indique cuáles serían las variables/niveles involucrados en su ejemplo y las hipótesis nula y alternativa a contrastar.

Buena suerte.