



## Ejercicio práctico nº 10

### Enunciado sala 2

#### 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
'anneal'         98.00      'cmc'         51.05
'contact-lenses' 68.33      'credit'      86.23
'ecoli'          80.04      'grub-damage' 47.79
'kr-s-kp'        92.46      'monks'       62.24
'monks1'         100.00     'mushroom'    95.83
'nursery'        94.28      'page-blocks' 93.51
'pasture-production' 85.83  'postoperatie' 66.67
'primary-tumor'  48.08      'segment'     91.30
'solar-flare-C'  88.24      'soybean'     92.08
'squash-stored'  58.00     'squash-unstored' 61.67
'tae'            44.38      'waveform'    79.86
'white-clover'   79.29      --            --
")
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  
'credit'     85.07   85.22   83.33  
'eucalyptus' 58.71   59.52   59.40  
'glass'      73.83   75.69   73.33  
'hepatitis'  83.79   82.50   81.25  
'iris'       92.67   92.00   93.33  
'optdigits'  96.90   94.20   91.80  
'page-blocks' 96.95   94.15   96.97  
'pendigits'  97.82   94.81   95.67  
'pima-diabetes' 75.01   74.75   72.67  
'primary-tumor' 47.49   49.55   38.31  
'solar-flare-C' 88.54   87.92   86.05  
'solar-flare-m' 87.92   86.99   85.46  
'solar-flare-X' 97.84   94.41   95.99  
'sonar'      81.26   80.79   78.36  
'waveform'   84.92   83.62   79.68  
'yeast'      57.74   57.48   56.26  
")  
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.**