



EJERCICIO PRÁCTICO 9: TRANSFORMACIÓN DE DATOS Y MÉTODOS NO PARAMÉTRICOS

Problema A

En trabajo de título de un estudiante del DIINF, se reportan los siguientes tiempos de ejecución (en milisegundos) medidos para dos versiones de un algoritmo genético para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

```
texto <- ("
Instancia 'Tiempo A6 (ms)' Instancia 'Tiempo B12 (ms)'
'f11400' 337977 'd1291' 335566
'pcb1173' 303634 'd657' 52696
'rat575' 33349 'f11577' 3192222
'rl1323' 243679 'nrw1379' 393213
'u1060' 3453176 'pr1002' 162808
'u1432' 398653 'pr2392' 8765321
'u1817' 876432 'rat783' 76857
'u2152' 3073534 'rl1304' 231254
'u574' 112326 'rl1889' 854213
'u724' 55026 'vm1084' 543215
")
```

Problema B

El siguiente texto muestra los resultados de clasificadores en diferentes instancias de prueba disponibles en el repositorio UCI. Los algoritmos corresponden a C1: J48 decision tree, C3: averaged one-dependence estimator (AODE), C5: J48 graft, C6: locally weighted naive-Bayes y C7: random forest. ¿Existe un algoritmo, o un grupo de algoritmos, mejor que otro(s)?

```
texto <- ("
Dataset C1 C3 C5 C6 C7
'credit' 84,93 85,07 84,93 85,22 83,33
'eucalyptus' 64,28 58,71 64,01 59,52 59,40
'glass' 71,58 73,83 71,10 75,69 73,33
'hepatitis' 79,46 83,79 79,46 82,50 81,25
'hungarian-14' 78,64 84,39 78,64 84,38 81,97
'hypothyroid' 99,28 98,54 99,28 98,62 98,97
'iris' 93,33 92,67 93,33 92,00 93,33
'mushroom' 100,00 99,95 100,00 99,84 100,00
'optdigits' 78,97 96,90 81,01 94,20 91,80
'page-blocks' 96,62 96,95 96,66 94,15 96,97
'pendigits' 89,05 97,82 89,87 94,81 95,67
'pima-diabetes' 73,70 75,01 73,56 74,75 72,67
'primary-tumor' 40,11 47,49 40,11 49,55 38,31
'solar-flare-C' 88,86 88,54 88,86 87,92 86,05
'solar-flare-m' 90,10 87,92 90,10 86,99 85,46
'solar-flare-X' 97,84 97,84 97,84 94,41 95,99
'sonar' 74,48 81,26 74,45 80,79 78,36
'waveform' 74,38 84,92 74,90 83,62 79,68
'yeast' 57,01 56,74 57,01 57,48 56,26")
```

Problema C

En trabajo de título de un estudiante del DIINF, se reportan los siguientes tiempos de ejecución ('Tpo' en milisegundos) medidos para dos versiones de un algoritmo genético (A6 y B12) para resolver instancias del problema del vendedor viajero disponibles en repositorios públicos. ¿Es uno de los algoritmos más rápido que el otro?

```
texto <- ("
Instancia 'Tpo A6' 'Tpo B12'
'rat575' 33349 32444
'u724' 55026 64019
'd657' 43352 52696
'rat783' 65076 76857
'u574' 112326 123456
'pr1002' 136262 162808
'fl1577' 3234574 3192222
'nrw1379' 335608 393213
'd1291' 268964 335566
'u1432' 398653 472597
'pcb1173' 303634 234658
'fl1400' 337977 430748
'u2152' 3073534 3253423
'rl1323' 243679 132654
'rl1304' 342321 231254
'u1817' 876432 672542
'vm1084' 413672 543215
'rl1889' 1876432 854213
'pr2392' 6764986 8765321
'u1060' 3453176 432876
")
```

Problema D

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 algoritmos 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 -- --
")
```

Problema E

Una compañía de cosméticos hizo una prueba preliminar de su nueva crema quitamanchas, en que 30 personas fueron separadas aleatoriamente en tres grupos de 10 voluntarios/as: uno de control, a quienes se les entregó una crema placebo (humectante solamente); otro que usaron la crema quitamanchas que la compañía comercializa actualmente; y el último que usaron el nuevo producto. A todos se les dijo que usaban la crema nueva de última generación. Dos personas del grupo de control y una del grupo con la crema existente abandonaron el estudio. Para el resto, se reportaron los siguientes números de manchas removidas al finalizar el tiempo de prueba:

```
texto <- ("
Nueva Actual Control
81 48 18
32 31 49
42 25 33
62 22 19
37 30 24
44 30 17
38 32 48
47 15 22
49 40 --
41 -- --
")
```

Problema F

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' 56,74 57,48 56,26
")
```