Licence d'Informatique 2ème année
AD Utilisateur (C5-160412-INFO)
1ère chance - 26 mars 2021 (1h45)

Carl FRÉLICOT – Dpt Info / Lab MIA

Exercice 1

L'objectif du *canvas* est la réduction de la dimension : $(178 \times 13) \rightarrow (9 \times 5)$ à partir d'un échantillon aléatoire ($n = 178 \times 5\% \approx 9$) stratifié [3,4,2]. Les $p = 5$ variables les plus discriminantes au sens de la statistique de Fisher sont sélectionnées.

Exercice 2

1) deux solutions acceptées
   - $Q_2 = 560$, $Q_1 = \frac{428+465}{2} = 446.5$, $Q_3 = \frac{770+1020}{2} = 895$ et $EIQ = 448.5$; moustaches : 425 et 1290, pas d'outlier
   - $Q_2 = 560$, $Q_1 = 465$, $Q_3 = 770$ et $EIQ = 305$; 425 et 1020, outlier en 1290
2) $s_B^2 = \frac{3}{9}(1026.67 - 681.44)^2 + \frac{4}{9}(517 - 681.44)^2 + \frac{2}{9}(492.5 - 681.44)^2 = 59\,678.88$ et
   $s_W^2 = \frac{3}{9}212.34^2 + \frac{4}{9}97.05^2 + \frac{2}{9}67.5^2 = 20\,228.01$; alors $F = \frac{59\,678.88/2}{20\,228.01/6} = 8.85$
3) $s_{CH}^2 = \frac{40.28}{9} - 5.044 \times 0.952 = -0.369$ et $r_{CH} = \frac{-0.369}{0.205 \times 1.975} = -0.811$;
   C et H sont fortement anti-corrélées et on doit pouvoir ajuster par un modèle affine décroissant
4) au sens de l'écart-type résiduel le modèle (3) est très légèrement meilleur que le (2) : $1.0243 < 1.0254$,
   mais le principe de parcimonie impose de retenir le modèle (2) plus simple : $C = a \times H^2 + b \times H + c$

Exercice 3

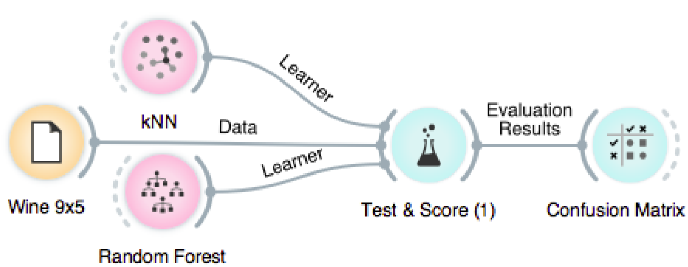
1) L'analyste a eu raison de choisir une ACP normée : $s_P^2 = 282.7^2 >> s_H^2 = 0.205^2$. Le % cumulé de variance expliquée suggère de retenir $q = 3$ ou 4 composantes (coude) et non 5.
   PC1 : axe colorimétrique, oppose les vins (1) et (2) à forte H et OD et faible C à gauche aux vins (3) à forte C et faibles H et OD à droite
   PC2 : axe (de force?) opposant les vins (1) à forts A et P à droite aux vins (2) à gauche à faibles A et P
2) On retrouve des éléments similaires à ceux de l'ACP.
   LDx : ordonne les vins par qualité : (3) à gauche (C+, OD- et H-), (2) au centre et (1) à droite (C-, OD+ et H+),
   LDy : oppose les vins (1) en haut (A+) aux vins (2) (A-) en bas
3) $\widehat{x}_x = {}^t x\, LD_x = [1, 600, 5, 11, 3][0.94, -0, -0.1, 0.15, 0.3] = 2.99$ et $\widehat{x}_y = {}^t x\, LD_y = [1, 600, 5, 11, 3][0.82, 0, 0.18, 0.54, -0.05] = 7.51$
4) $d^2(\widehat{x}, \overline{x}_3) = {}^t(\widehat{x} - \overline{x}_3)(\widehat{x} - \overline{x}_3) = [0.89, -2.67][0.89, -2.67] = 7.921$

Exercice 4

1) $t_{13} = \frac{3 \times 1}{9} = \frac{1}{3}$ et $e_{13} = \frac{(1-/3)^2}{1/3} = \frac{4}{3}$, soit $\frac{4/3}{9} = 14.81\%$ du χ^2 d'indépendance
2) la p-value vaut 6.01%, un risque assez faible de rejeter à tort l'hypothèse d'indépendance entre *Cluster* et *Wine* ; elles sont donc plutôt dépendantes et *Cluster* est plutôt conforme à la vérité-terrain

Exercice 5

1)



2) confusion (1) :

2	1
1	5

alors $R_1 = \frac{2}{2+1} = 2/3 = 0.67$

confusion (2) :

2	2
2	3

alors $P_2 = \frac{2}{2+3} = 2/5 = 0.4$

3) comme $P_3 = 0$, alors $P = \frac{3 \times \frac{2}{3} + 3 \times \frac{2}{5} + 2 \times 0}{9} = 0.40$