

- Autorisés : calculatrice (pas une application sur smartphone), un formulaire **manuscrit** sans **aucun exemple numérique**. Tout **autre document est interdit**
- Utilisez les cadres réservés pour inscrire vos réponses.  
Vous devez **écrire la formule** utilisée avant de donner le détail du moindre calcul.
- Une réponse **non justifiée** sera considérée comme **fausse**.

NOM (en **capitales**), Prénom :

Note : /20

Le tableau ci-contre donne les notes cumulées attribuées par 10 usagers à 8 smartphones pour les caractéristiques suivantes :

- sonnerie (Bell) notée de 1 à 3 (faible, moyen, fort)
- qualité Sonore (Sound) de 1 (exécrable) à 10 (excellent)
- qualité de Transmission (Trans.) de 1 à 10
- autonomie (Auto.) de 1 à 10
- facilité d'Utilisation (Use) de 1 à 10

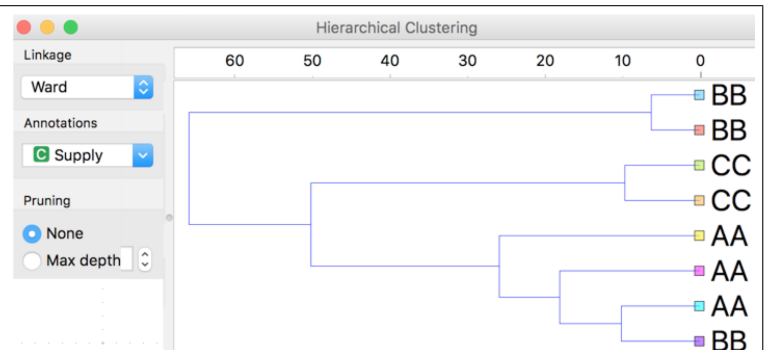
On dispose aussi pour chaque smartphone de l'Avis d'un site web de test et d'un code du fabricant (Supply).

ID.	Bell	Sound	Trans.	Auto.	Use	Avis	Supply
#1	21.0	67.0	88.0	75.0	60.0	1 étoile	AA
#2	19.0	58.0	61.0	40.0	70.0	5 étoiles	BB
#3	20.0	62.0	78.0	69.0	68.0	3 étoiles	AA
#4	14.0	58.0	54.0	86.0	62.0	1 étoile	CC
#5	20.0	66.0	70.0	63.0	79.0	5 étoiles	AA
#6	17.0	56.0	56.0	93.0	67.0	1 étoile	CC
#7	17.0	58.0	56.0	40.0	73.0	5 étoiles	BB
#8	20.0	59.0	69.0	70.0	65.0	3 étoiles	BB

## Exercice 1

La sortie Orange d'un clustering hiérarchique, avec la distance euclidienne usuelle, est donnée ci-dessous.

- 1) Les 2 premiers smartphones regroupés sont #2 et #7. Calculez quelle devrait être la valeur de l'indice correspondant pour qu'il s'agisse bien de l'accroissement d'inertie intra-groupes.



- 2) Quelle partition retiendriez-vous, et pourquoi ?

- 3) Complétez le canevas ci-après dont le but sera de réaliser ce clustering, puis de comparer la partition résultante à la variable catégorielle Supply.



## Exercice 2

Pour alléger les calculs, on a réduit le tableau de données à 3 variables, transposé ci-dessous à gauche.

ID.	#1	#2	#3	#4	#5	#6	#7	#8	
Supply	AA	BB	AA	CC	AA	CC	BB	BB	
Bell	21.000	19.000	20.000	14.000	20.000	17.000	17.000	20.000	$d_\infty$ : #1 #2 #3 #4 #5 #6 #7 #8
Auto.	75.000	40.000	69.000	86.000	63.000	93.000	40.000	70.000	$x$ 39 41 40 46 40 43 30 40
Use	60.000	70.000	68.000	62.000	79.000	67.000	73.000	65.000	$y$ 10 30 2 26 10 24 30 11
									$z$ 10 30 3 16 9 23 30 5

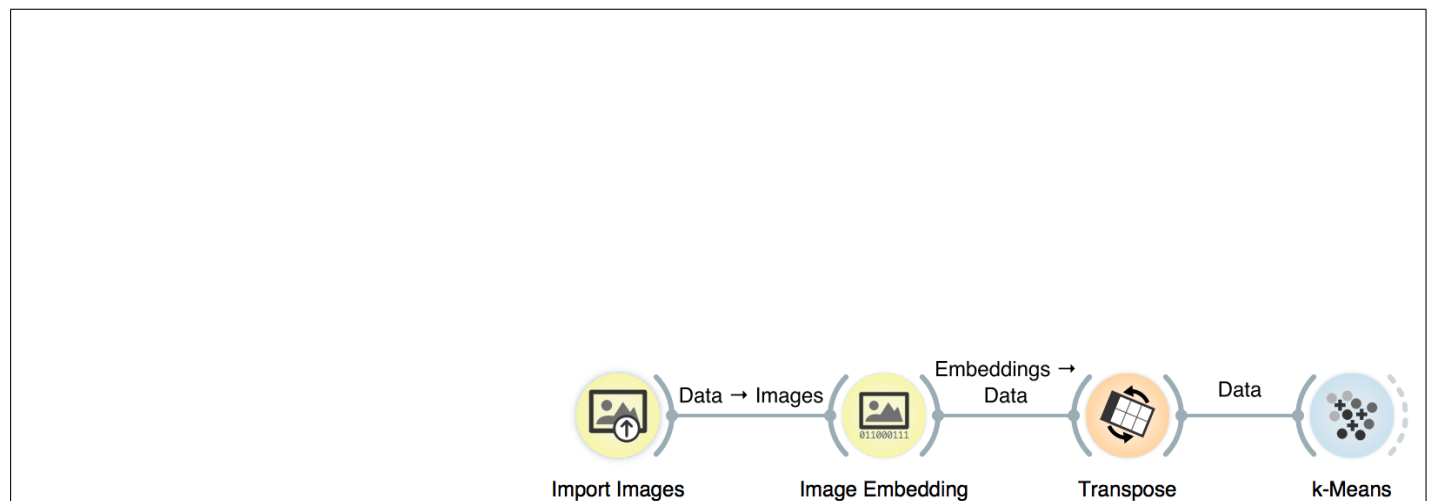
On souhaite classer les smartphones  $x = (60, 50, 75)$ ,  $y = (17, 70, 70)$  et  $z = (18, 80, 65)$  à l'aide de la règle des 3-*Nearest Neighbors*. Pour cela, on donne ci-dessus à droite leur distance de *Chebychev* aux données d'apprentissage.

- 1) Calculez la distance manquante.

- 2) Classez  $x$ ,  $y$  et  $z$ . Vous donnerez la liste des voisins, leur groupe, les probabilités de classement, et enfin le classement.

## Exercice 3

Expliquez brièvement ce que pourrait-être le but final du canevas ci dessous pour un analyste.



Exercice 4

On considère le même sous-tableau de données qu'à l'Exercice 2 sur lequel on a exécuté deux itérations des *K-Means* avec  $K = 3$  et la distance euclidienne usuelle.

1) Faites-les calculs permettant de donner les valeurs manquantes dans les tableaux ci-dessous.

$$Y^{(0)} = [3, 2, 1, 1, 2, 3, 3, 2] \rightarrow V^{(1)} = \begin{bmatrix} \bar{x}_1 = ( \quad , \quad , \quad ) \\ \bar{x}_2 = (19.67, 57.67, 71.33) \\ \bar{x}_3 = (18.33, 69.33, 66.67) \end{bmatrix} \rightarrow$$

$d_2^2$	#1	#2	#3	#4	#5	#6	#7	#8
$\bar{x}_1$	47.25	1435.25		90.25	415.25	244.25	1470.25	65.25
$\bar{x}_2$	430.67	314.33	139.67	922	87.33	1274.33	322	192.33
$\bar{x}_3$	83.67	872	4.67	318.33	195	562	902.33	6
$Y^{(1)}$	1	2	3	1	2	1	2	3

$$\rightarrow V^{(2)} = \begin{bmatrix} \bar{x}_1 = (17.33, 84.67, 63) \\ \bar{x}_2 = ( \quad , \quad , \quad ) \\ \bar{x}_3 = (20, 69.5, 66.5) \end{bmatrix} \rightarrow$$

$d_2^2$	#1	#2	#3	#4	#5	#6	#7	#8
$\bar{x}_1$	115.89	2046.56	277.56	13.89	732.56	85.56	2095.22	226.22
$\bar{x}_2$	948.56	74.89	492.89	1635.22	261.89	2106.89	62.56	581.56
$\bar{x}_3$	73.5	883.5	2.5	328.5	198.5		921.5	2.5
$Y^{(2)}$	3	2	3	1	3	1	2	

2) Calculez le plus simplement possible l'inertie intra-groupes de la partition  $Y^{(2)}$ .

3) L'inertie intra-groupes de la partition Supply est vaut 66.26 ; est-elle meilleure que  $Y^{(2)}$  ?

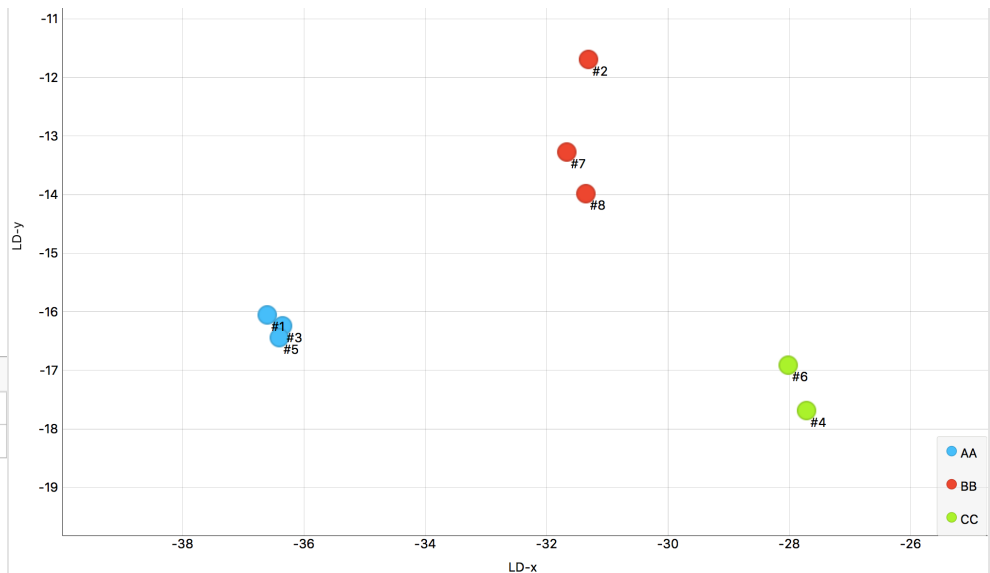
4) Calculez la distance cosinus entre Bell et Auto.

### Exercice 5 :

On réalise, sur le tableau initial (page 1), une Analyse Factorielle Discriminante avec Orange, où les groupes sont donnés par Supply.

Le 1er plan discriminant est visualisé ci-contre, les vecteurs directeurs sont donnés ci-dessous, ainsi que les centroids des groupes :

component	Bell	Sound	Trans.	Auto.	Use
LD-x	0.80	0.10	-0.46	0.03	-0.36
LD-y	0.95	0.00	-0.19	-0.09	-0.21
$\bar{x}_{AA}$	-36.4	-16.2			
$\bar{x}_{BB}$	-31.4	-13.0			
$\bar{x}_{CC}$	-27.9	-17.3			



- 1) Pourrait/devrait-on regarder une troisième variable LD-z ?

On souhaite classer un nouveau smartphone  $x = (20, 60, 50, 70, 80)$  dans le plan discriminant 1-2 à l'aide de la règle du *Nearest Prototype* au sens de la distance de *Manhattan*.

- 2) Sa projection sur LD-x vaut -27.7 ; calculez celle sur LD-y.

- 3) On donne  $d_1(x, \bar{x}_{AA}) = 11.3$  et  $d_1(x, \bar{x}_{BB}) = 4.3$ . Faites le reste des calculs nécessaires prédire quel fabricant (Supply) associer à ce nouveau smartphone. N'oubliez pas de conclure.

- 4) Visuellement, aurait-on la même prédiction avec la règle du *Nearest Neighbor* ? Vous pouvez utiliser la figure et l'annoter, et/ou expliquer brièvement.