Exercise 1 ACP des images 13 points

Le but de cet exercice est d'analyser globalement le tableau de données x qui contient 2007 imagettes en niveau de gris (normalisé entre 0 et 2) de taille 16×16, chaque imagette 16×16 ayant été transformée en un vecteur de 256 caractéristiques en mettant bout à bout toutes ses colonnes. Associé à chaque imagette, on dispose d'une étiquette stockée dans le vecteur $y \in \{1, \dots, 10\}^{2007}$ qui code de 1 à 10 le chiffre représenté par l'imagette (les zéros sont codés par 10). Les données sont disponibles dans le fichier `median_m8_2017.mat`, des données du médian 2017 dans l'onglet Annales du Moodle de M8. Les trois premières questions ne seront pas notées.

1. Chargez les données et vérifiez que la matrice x et le vecteur y ont la bonne taille

```
load median_m8_2017
whos
```

2. Visualisez la troisième imagette et vérifiez qu'il s'agit bien d'un trois.

```
image = reshape(x(3,:),16,16);
figure(1)
colormap gray
imagesc(image')
y(3)
```

3. Représentez les 150 ème et 200 ème variables des un, les neufs et les zéros sur un même graphique avec des couleurs différentes. Qu'en pensez vous ?

```
figure(2)
hold on
ind1 = find(y==1);
plot(x(ind1,150),x(ind1,200),'ro')
ind9 = find(y==9);
plot(x(ind9,150),x(ind9,200),'gx')
ind0 = find(y==10);
plot(x(ind0,150),x(ind0,200),'c*')
```

4. Calculez les 12 premiers axes factoriels du tableau x
5. Quel est le nombre minimal d'axes factoriels qu'il faut considérer pour représenter au moins 50% de l'information contenue dans le tableau x ?
6. Les trois premières composantes principales.
   a) Visualisez sur les deux premières composantes principales, les individus correspondant à des un, les neufs et les zéros, sur un même graphique avec des couleurs différentes. Faite la même figure avec les première et troisième composantes, puis avec les deuxième et troisième composantes.
   b) Quelles sont, après votre analyse des figures des composantes principales, les deux qui permettraient le mieux de distinguer des imagettes de un, de neufs et de zéros ?
7. Comment interpréter les axes en terme de variables ?
   a) Calculez les corrélations entre données et les composantes principales.
   b) Visualisez les 4 premiers axes factoriels.
   c) Commentez les.

Exercise 2 **7 points**

Après avoir acheté 10 cerises une par une sur un marché, un acteur qui joue dans des publicités les a toutes pesées et, avant de les manger, a reporté ses résultats dans le tableau suivant :

Cerise	1	2	3	4	5	6	7	8	9	10
Poids (en g)	2.51	3.03	2.90	3.01	3.26	3.25	2.92	2.97	3.08	3.07

1. Calculez la moyenne, la médiane et la variance des poids des cerises sans utiliser les fonctions matlab `mean, median, var et std`.
2. Vérifiez que vos calculs sont exacts.
3. Soit n le nombre d'observations (ici $n = 10$). En notant $x_i, i = 1, \ldots, n$ le poids de chaque cerise, on se donne la fonction suivante d'un paramètre v.

$$J(v) = \ln \left[\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi v}} \exp^{-\frac{(x_i-3)^2}{2v}} \right) \right].$$

On pourra commencer par développer $J(v)$ en utilisant le formulaire ci-dessous. Montrez que, en supposant $\frac{1}{n} \sum_{i=1}^{n} x_i = 3$, la valeur $\hat{v}$ du paramètre v qui minimise $J(v)$ s'écrit :

$$\hat{v} = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - 9.$$

4. Calculez $\hat{v}$ et commentez.

Formulaire

- $\ln(ab) = \ln a + \ln b$
- $\ln \frac{a}{b} = \ln a - \ln b$
- $\ln a^b = b \ln a$
- $\ln(\exp(a)) = a$