Exam

Exercise 1

Based on your reading of the article "Deep Dictionary Learning" by S. Tariyal *et al.*, answer the following questions :

1. Why do the authors say that "the dictionary learning formulation in equation (2) is **unsupervised**" ?
2. Recall the definition of $\|Z\|_0$ involved in (2). What is the goal for imposing such constraint ?
3. Give the ISTA algorithm, when applied to the resolution of Problem (5) with respect to variable Z. Precise the condition on the stepsize, that ensures convergence of the method.
4. What do the authors mean by "Here we propose to learn the dictionaries in a **greedy manner**" ?

Exercise 2

Based on your reading of the article "Matrix completion incorporating auxiliary information for recommender system design" by A. Gogna *et al.*, answer the following questions :

1. Give an advantage of the low rank matrix completion approach, with respect to the matrix factorization formulation ?
2. Explain the procedure for applying NMF to movie recommendation when each user has rated the movies he/she has seen from 1 to 5. How would you measure the performance of the model ?
3. Explain the role of μ_G and W_G, in equation (15).
4. What is the (closed form) solution to (15) when $\lambda = 0$?

Exercise 3

Let us consider a mixture of K real-valued generalized Gaussian distributions whose PDF is given by $f(y) = \sum_{i=1}^{K} w_i \, f_{\theta_i}(y)$, with

$$f_\theta(y) = \frac{\beta}{2\,\alpha\,\Gamma(1/\beta)} \exp\left(-(|y-\mu|/\alpha)^\beta\right)$$

where $\theta = (\mu, \alpha, \beta) \in \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*^+$ (μ is the location parameter, α the scale and β the shape) and $\Gamma(t) = \int_0^\infty x^{t-1} \exp(x)dx$ is the gamma function.

1. Which value a of β provides Gaussian distributions ?
2. Discuss the different shape of the distribution according to β and motivate the interest of such distributions.

First analysis : K and β_i's known

3. Given an independent and identically distributed n-sample $(x_1, \ldots, x_n)$ of PDF $f_{\mu,\alpha}(.)$, derive the Maximum Likelihood estimators for (μ, α).
4. Let us now consider n independent observations $(y_1, \ldots, y_n)$ distributed according to the mixture $f(.)$, propose the EM algorithm to estimate (w_i, μ_i, α_i), for $i = 1, \ldots, K$.

Second analysis : K known, β_i's unknown

5. Propose a way for estimating β given an independent and identically distributed n-sample $(x_1, \ldots, x_n)$ of PDF f_θ.
   Indication : it is an optimization problem... first, you may assume that μ and α are known. You may also use the digamma function, defined as $\psi(x) = \frac{d}{dx} \log(\Gamma(x)) = \dfrac{\Gamma(x)}{\Gamma'(x)}$ and the trigamma function that is simply $\psi'(x)$.
6. Propose the general EM algorithm for a mixture of K real-valued generalized Gaussian distributions.

Third Analysis : K unknown

7. Propose an approach to estimate K in this context of mixture of K real-valued generalized Gaussian distributions.

8. Which clustering algorithms would you try for each of the following situations :
   - Dataset with outliers
   - Clusters with different densities
   - Round-shaped clusters
   - Non-convex clusters
   - Overlapped ellipsoid clusters