Problem Set: Power k-Means Clustering

The paper introduces a novel algorithm for the classical task of k-means clustering, the power k-means algorithm. The problem set is designed to facilitate the understanding of power k-means clustering and test its performance on additional data set. You will show the derivation process of the formulas discussed in the paper and implement the algorithm by hand.

1. Define the power mean of a vector y:

$$M_s(y) = \left(\frac{1}{k} \sum_{i=1}^{k} y_i^s \right)^{\frac{1}{s}}$$

   where $y \in (0, \infty)$. Explore the concavity and convexity of $M_s(y)$.
   Hint: Examine the convexity by computing the gradient and Hessian matrix of $M_s(y)$. Determine the condition on s, under which $M_s(y)$ is convex or concave on the set $y \in (0, \infty)^k$.

2. Define the power k-means objective function for given powers s:

$$f_s(\Theta) = \sum_{i=1}^{n} M_s(||x_i - \theta_1||^2, ..., ||x_i - \theta_k||^2).$$

   Given the linear majorization derived from the concavity of $M_s(y)$:

$$M_s(y) \leq M_s(y_m) + \sum_{j=1}^{K} \nabla_y M_s(y_m)^T (y_j - y_{m,j}).$$

   Define the right-hand side of the inequality as the majorizing surrogate function. Construct MM step in the power k-means clustering by solving the weights of the majorization $\omega_{m,ij}$ and the update rule $\theta_{m+1,j}$ that minimizes the weights. Hint: Focus only on the terms on the right-hand side that is relevent to Θ when minimizing the majorization.

3. Implement power k-means algorithm (Algorithm 2) and test your implementation on the Fashion-MNIST data set. Report NMI score on the test and train data sets: see `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.normalized_mutual_info_score.html`. Additionally, set $s_0 = -2$ and $\eta = 1.05$ by default, and try different selection for the initial power s_0 and incremental factor *eta*. Observe the performance. Choose the parameters based on performance or explain why the initial value s_0 do not require careful tuning.