CS 395T **Homework 2** **Due: Wed. Nov. 5, 2003**

1. The *spherical k-means* algorithm uses cosine similarity for clustering. The corresponding objective function that it tries to maximize is

$$\sum_{j=1}^{k} \sum_{\mathbf{x} \in \mathcal{C}_j} \mathbf{x}^T \mathbf{c}_j, \quad \text{with } \|\mathbf{x}\| = 1, \ \|\mathbf{c}_j\| = 1, \tag{1}$$

where $\mathbf{c}_j$ is the L_2-normalized mean vector of the j-th cluster, $\mathcal{C}_j$.

   a Show that maximizing (1) is equivalent to minimizing

$$\sum_{j=1}^{k} \sum_{\mathbf{x} \in \mathcal{C}_j} \|\mathbf{x} - \mathbf{c}_j\|^2, \quad \text{with } \|\mathbf{x}\| = 1, \ \|\mathbf{c}_j\| = 1. \tag{2}$$

   b Give pseudocode for the spherical k-means algorithm.

   c Show that this algorithm monotonically increases the objective function in (1) (or monotonically decreases the objective function in (2)).

2. Consider a strictly convex differentiable function $\phi : S \to R$, where S is a convex set, $S \subseteq R^d$. The *Bregman divergence* corresponding to ϕ, $D_\phi : S \times S \to R$ is defined as

$$D_\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) - \phi(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^T \nabla\phi(\mathbf{y}),$$

where $\nabla\phi(\mathbf{y})$ is the gradient of ϕ at $\mathbf{y}$. Due to the convexity of ϕ, it can be shown that $D_\phi(\mathbf{x}, \mathbf{y}) \geq 0$ for all x, y.

   a Consider a set of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in S$. Show that the mean $\mathbf{m} = \frac{1}{n} \sum_i \mathbf{x}_i$ is the best "Bregman" representative of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, i.e.,

$$\sum_{i=1}^{n} D_\phi(\mathbf{x}_i, \mathbf{m}) \leq \sum_{i=1}^{n} D_\phi(\mathbf{x}_i, \mathbf{z}) \quad \text{for any } \mathbf{z} \in R^d.$$

   b Calculate $D_\phi(\mathbf{x}, \mathbf{y})$ when $S = R^d$ and $\phi(\mathbf{x}) = \|\mathbf{x}\|_2^2$.

   c Calculate $D_\phi(\mathbf{p}, \mathbf{q})$ when S is the unit simplex, i.e., $S = \{\mathbf{p} | \ \mathbf{p} \geq 0$ and $\sum_{j=1}^{d} p_j = 1\}$, and $\phi(\mathbf{p}) = \sum_i p_i \log p_i$ (log is to the natural base).

   d Based on fact a above, give pseudocode of a k-means algorithm that uses Bregman divergences. What is the global objective function that your algorithm monotonically decreases?