CS 395T # Homework 3 **Due: Tue, Nov 18, 2004**

1. Consider minimizing the following objective function for clustering:

$$\sum_{j=1}^{k} \sum_{\mathbf{x} \in \mathcal{C}_j} \|\mathbf{x} - \mathbf{c}_j\|_1, \tag{1}$$

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

   a How would you define the representative vector for a cluster? Is there a quick way to compute this vector?

   b Give pseudocode for a k-means like algorithm that tries to minimize the above objective function.

   c Does this algorithm monotonically decrease the above objective function?

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 $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, where $\mathbf{A}$ is a positive definite matrix.

   c Calculate $D_\phi(\mathbf{p}, \mathbf{q})$ when S is the positive orthant, i.e., $S = \{\mathbf{x} | \ \mathbf{x} \geq 0\}$, and $\phi(\mathbf{p}) = -\sum_i \log p_i$ (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? Can the objective function in 1 be considered a special case?