Problem 1. k-Means Clustering.

Show that the k-means algorithm might converge to a point that is not a local minimum.

Hint: Suppose that $k = 2$ and the sample points are $\{1, 2, 3, 4\} \subset \mathbb{R}$. Suppose we initialize the k-means with the centers $\mu_1 = 2, \mu_2 = 4$; and suppose we break ties in the definition of C_i by assigning i to be the smallest value in $\arg\min_j \|\mathbf{x} - \mu_j\|$.

Solution.

Problem 2. Spectral Clustering.

In lecture, we studied spectral partitioning using the *Fiedler vector*, an eigenvector corresponding to the second-smallest eigenvalue of the graph Laplacian. In this problem, we analyze a basic algorithm for computing a leading eigenvector: the *power method*.

1. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite matrix with eigenvalues

$$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

and corresponding orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Starting from any vector $\mathbf{x}^0 = \sum_{i=1}^n a_i \mathbf{v}_i$ with $a_1 \neq 0$, the power method performs the iterations

$$\mathbf{x}^{t+1} = \frac{A\mathbf{x}^t}{\|A\mathbf{x}^t\|_2}, \qquad t = 0, 1, \ldots.$$

Prove that, for every $i \in \{2, \ldots, n\}$ and every integer $t \geq 1$,

$$\left| \frac{\langle \mathbf{x}^t, \mathbf{v}_i \rangle}{\langle \mathbf{x}^t, \mathbf{v}_1 \rangle} \right| \leq \left(\frac{\lambda_i}{\lambda_1} \right)^t \left| \frac{\langle \mathbf{x}^0, \mathbf{v}_i \rangle}{\langle \mathbf{x}^0, \mathbf{v}_1 \rangle} \right|.$$

Thus, $\mathbf{x}^t$ converges to the top eigenvector $\pm\mathbf{v}_1$.

2. Let $\alpha = a_1^2 / \|\mathbf{x}^0\|_2^2$. Prove that, for every integer $t \geq 1$,

$$(\mathbf{x}^t)^\top A \mathbf{x}^t \geq \lambda_1 \alpha^{1/(2t+1)}.$$

That is, for any $\varepsilon \in (0, 1)$, every integer $t \geq \max\{1, \frac{1}{2\varepsilon} \log(1/\alpha)\}$ satisfies

$$(\mathbf{x}^t)^\top A \mathbf{x}^t \geq (1 - \varepsilon)\lambda_1.$$

Unlike convergence to the leading eigenvector, this guarantee does not depend on the spectral gap $\lambda_1 - \lambda_2$.

Hint: Apply Jensen's inequality to the concave function $u \mapsto u^{2t/(2t+1)}$.

Solution.