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.

For centers $(\mu_1, \mu_2) \in \mathbb{R}^2$, define the k-means objective

$$G(\mu_1, \mu_2) = \sum_{x \in \{1,2,3,4\}} \min\{(x - \mu_1)^2, (x - \mu_2)^2\}.$$

At these centers, points 1 and 2 are strictly closer to $\mu_1 = 2$, and point 4 is strictly closer to $\mu_2 = 4$. Point 3 has squared distance 1 from each center, so the specified tie rule assigns it to cluster 1. Thus the assignment step gives

$$C_1 = \{1, 2, 3\}, \qquad C_2 = \{4\}.$$

The update step replaces each center by the mean of its cluster, giving

$$\mu_1^{\text{new}} = \frac{1 + 2 + 3}{3} = 2, \qquad \mu_2^{\text{new}} = 4.$$

Thus, every subsequent iteration makes the same assignments and leaves both centers unchanged. The algorithm therefore converges to $(2, 4)$, where

$$G(2, 4) = 1 + 0 + 1 + 0 = 2.$$

For $0 < \delta < 1$, consider the centers $(2, 4 - \delta)$. Their midpoint is $3 - \delta/2 \in (5/2, 3)$. Since $2 < 4 - \delta$, points $1, 2$ are strictly closer to the first center, and points $3, 4$ are strictly closer to the second. Hence

$$\begin{aligned} G(2, 4 - \delta) &= (1 - 2)^2 + (2 - 2)^2 + \big(3 - (4 - \delta)\big)^2 + \big(4 - (4 - \delta)\big)^2 \\ &= 2 - 2\delta + 2\delta^2 = 2 - 2\delta(1 - \delta) < 2. \end{aligned}$$

The distance from $(2, 4 - \delta)$ to $(2, 4)$ in the center parameter space $\mathbb{R}^2$ is exactly δ. Given any radius $r > 0$, choosing $0 < \delta < \min\{1, r\}$ therefore produces a point within that radius whose objective is strictly smaller, which means $(2, 4)$ is not a local minimum.

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.

1. For every integer $t \geq 1$, we have

$$A^t \mathbf{x}^0 = \sum_{j=1}^n a_j \lambda_j^t \mathbf{v}_j, \qquad D_t := \|A^t \mathbf{x}^0\|_2 = \left(\sum_{j=1}^n a_j^2 \lambda_j^{2t} \right)^{1/2} > 0.$$

We claim that the iterations satisfy

$$\mathbf{x}^t = \frac{A^t \mathbf{x}^0}{D_t} = \frac{\sum_{j=1}^n a_j \lambda_j^t \mathbf{v}_j}{D_t}, \qquad t \geq 1. \tag{1}$$

For $t = 1$, this is the defining update, whose denominator D_1 is nonzero. If it holds at time t, then

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

Normalizing proves the formula at time $t+1$.

Taking inner products in (1) yields

$$\langle \mathbf{x}^t, \mathbf{v}_i \rangle = \frac{a_i \lambda_i^t}{D_t}, \qquad \langle \mathbf{x}^t, \mathbf{v}_1 \rangle = \frac{a_1 \lambda_1^t}{D_t} \neq 0.$$

Thus for every $i \geq 2$ and $t \geq 1$ we have

$$\left| \frac{\langle \mathbf{x}^t, \mathbf{v}_i \rangle}{\langle \mathbf{x}^t, \mathbf{v}_1 \rangle} \right| = \left(\frac{\lambda_i}{\lambda_1} \right)^t \left| \frac{a_i}{a_1} \right| = \left(\frac{\lambda_i}{\lambda_1} \right)^t \left| \frac{\langle \mathbf{x}^0, \mathbf{v}_i \rangle}{\langle \mathbf{x}^0, \mathbf{v}_1 \rangle} \right|.$$

To verify convergence of the vectors themselves, factor $|a_1|\lambda_1^t$ from the numerator and denominator of (1). This gives

$$\mathbf{x}^t = \frac{\dfrac{a_1}{|a_1|}\mathbf{v}_1 + \displaystyle\sum_{i=2}^{n} \frac{a_i}{|a_1|} \left(\frac{\lambda_i}{\lambda_1} \right)^t \mathbf{v}_i}{\left[1 + \displaystyle\sum_{i=2}^{n} \frac{a_i^2}{a_1^2} \left(\frac{\lambda_i}{\lambda_1} \right)^{2t} \right]^{1/2}}.$$

For every $i \geq 2$, we have $0 \leq \lambda_i/\lambda_1 < 1$. Each corresponding power tends to zero. Since there are only finitely many terms, the numerator converges in Euclidean norm to $(a_1/|a_1|)\mathbf{v}_1$ and the denominator converges to 1. Therefore

$$\lim_{t \to \infty} \mathbf{x}^t = \frac{a_1}{|a_1|}\mathbf{v}_1 = \begin{cases} \mathbf{v}_1, & a_1 > 0, \\ -\mathbf{v}_1, & a_1 < 0. \end{cases}$$

2. Define weights and normalized eigenvalues by

$$p_i := \frac{a_i^2}{\|\mathbf{x}^0\|_2^2}, \qquad r_i := \frac{\lambda_i}{\lambda_1}, \qquad i = 1, \ldots, n.$$

Because the eigenvectors form an orthonormal basis, $\|\mathbf{x}^0\|_2^2 = \sum_i a_i^2 > 0$. Hence,

$$p_i \geq 0, \qquad \sum_{i=1}^{n} p_i = 1, \qquad p_1 = \alpha \in (0, 1], \qquad r_1 = 1, \quad 0 \leq r_i \leq 1.$$

For each real $s > 0$, write

$$S_s := \sum_{i=1}^{n} p_i r_i^s.$$

The first summand is $p_1 r_1^s = \alpha$, so $S_s \geq \alpha > 0$. Using (1) and orthonormality, we obtain

$$(\mathbf{x}^t)^\top A \mathbf{x}^t = \frac{\sum_{i=1}^{n} a_i^2 \lambda_i^{2t+1}}{\sum_{i=1}^{n} a_i^2 \lambda_i^{2t}} = \lambda_1 \frac{S_{2t+1}}{S_{2t}}. \tag{2}$$

Fix $t \geq 1$ and let $q = 2t/(2t+1) \in (0,1)$. Since $f(u) = u^q$ is concave on $[0,\infty)$, by Jensen's inequality applied with weights p_i and arguments r_i^{2t+1}, we get that

$$S_{2t} = \sum_{i=1}^{n} p_i \big(r_i^{2t+1}\big)^q \leq \left(\sum_{i=1}^{n} p_i r_i^{2t+1} \right)^q = S_{2t+1}^{2t/(2t+1)}.$$

Combining this inequality with (2) and $S_{2t+1} \geq \alpha$ proves

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

Finally, for t satisfying

$$t \geq \max\left\{1, \frac{\log(1/\alpha)}{2\varepsilon}\right\},$$

we have

$$0 \leq \frac{\log(1/\alpha)}{2t+1} \leq \frac{\log(1/\alpha)}{2t} \leq \varepsilon.$$

Using $e^{-x} \geq 1 - x$, we have

$$\alpha^{1/(2t+1)} = \exp\left(-\frac{\log(1/\alpha)}{2t+1}\right) \geq 1 - \frac{\log(1/\alpha)}{2t+1} \geq 1 - \varepsilon.$$

It follows that

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