Mathematics of Machine Learning Homework 1- Solutions

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

March 5, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously.

Problem 1

Given a set of points $x_1, \dots, x_n \in \mathbb{R}^d$ and a partition of them into k clusters $S_1, \dots, S_k$, recall that the k-means objective is

$$\min_{S_1,\dots,S_k} \min_{\mu_1,\dots,\mu_k} \sum_{l=1}^{k} \sum_{i \in S_l} \|x_i - \mu_l\|_2^2.$$

If $\hat{S}_1, \dots, \hat{S}_k$ are the minimizers for the k-means objective, show that they are also minimizers for the optimization problem below

$$\min_{S_1,\dots,S_k} \sum_{l=1}^{k} \frac{1}{|S_l|} \sum_{i,j \in S_l} \|x_i - x_j\|_2^2.$$

Solution 1

We claim that for every collection of points $y_1, \dots, y_m \in \mathbb{R}^d$, the function $f(\mu) := \sum_{i=1}^m \|y_i - \mu\|_2^2$ is minimized when $\mu = \frac{1}{m}\sum_{i=1}^m y_i$. Observe that the function is convex so first order condition is enough to find the global minimum, we write

$$\nabla_\mu \sum_{i=1}^m \|y_i - \mu\|_2^2 = 0 \longrightarrow \sum_{i=1}^m -2(y_i - \mu) = 0 \longrightarrow \mu = \frac{1}{m}\sum_{i=1}^m y_i.$$

By the claim above, $\mu_l = \frac{1}{|S_l|}\sum_{i\in S_l} x_i$. Now we expand terms

$$\begin{aligned}\sum_{i,j\in S_l} \|x_i - x_j\|_2^2 &= \sum_{i,j\in S_l} \left(\|x_i\|_2^2 + \|x_j\|_2^2 - 2\langle x_i, x_j\rangle\right) \\ &= \sum_{i\in S_l} \left(|S_l|\|x_i\|_2^2 + \sum_{j\in S_l}\|x_j\|_2^2 - 2|S_l|\langle x_i, \mu_l\rangle\right) \\ &= 2|S_l|\sum_{i\in S_l}\|x_i\|_2^2 - 2|S_l|^2\|\mu_l\|_2^2.\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_{i\in S_l} \|x_i - \mu_l\|_2^2 &= \sum_{i\in S_l} \|x_i\|_2^2 + |S_l|\|\mu_l\|_2^2 - 2|S_l|\|\mu\|_2^2 \\ &= \sum_{i\in S_l}\|x_i\|_2^2 - |S_l|\|\mu_l\|_2^2.\end{aligned}$$

Therefore $\sum_{i\in S_l}\|x_i - \mu_l\|_2^2 = \frac{1}{2|S_l|}\sum_{i,j\in S_l}\|x_i - x_j\|_2^2.$. Summing over $l \in \{1, \dots, k\}$ and taking the minimum over $S_1, \dots, S_k$ finishes the proof.

Problem 2

Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ its eigenvalues. Describe the solutions of the optimization problems below

(a)

$$\max_{x\in\mathbb{S}^{n-1}} \langle Ax, x\rangle.$$

(b)

$$\max_{x\in\mathbb{S}^{n-1}} \|Ax\|_2.$$

Solution 2

(a) By the standard spectral theorem, $A = Q\Lambda Q^T$ where Q is an orthogonal matrix and Λ is a diagonal matrix. Then, for all unit vectors x,

$$\begin{aligned}\langle Ax, x\rangle &= \langle \Lambda Q^T x, Q^T x\rangle \\ &= \langle \Lambda y, y\rangle \; (y := Q^T x) \\ &= \sum_{i=1}^{n} \lambda_i (y_i)^2 \\ &\leq \lambda_1 \sum_{i=1}^{n} (y_i)^2 = \lambda_1.\end{aligned}$$

The last step follows from the fact that $\|y\|_2 := \|Q^T x\|_2 = \|x\|_2 = 1$. Now it is clear that the normalized eigenvector associated with the eigenvalue λ_1 maximizes the quadratic form and its maximum is λ_1.

(b) Observe that $\|Ax\|_2^2 = \langle Ax, Ax\rangle = \langle x, A^T Ax\rangle = \langle x, A^2 x\rangle$, therefore by letter "a" we obtain that $\|Ax\|_2^2 \leq \max\{\lambda_1^2, \lambda_n^2\}$.

Now it is clear that

$$\max_{x \in S^{n-1}} \|Ax\|_2 = \max\{|\lambda_1|, |\lambda_n|\},$$

because both values can be attained by the corresponding normalized eigenvectors v_1 and v_n.