University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Solution 8

Problem 1: Prove the following Courant-Fischer Min-Max Characterisation of Eigenvalues. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and corresponding eigenvectors $f_1, \ldots, f_n$. Then, it holds for any $1 \leq i \leq n$ that

$$\lambda_i = \min_{S:\dim S=i} \quad \max_{x\in S, x\neq 0} \frac{x^\intercal Ax}{x^\intercal x} = \max_{S:\dim S=n-i+1} \quad \min_{x\in S, x\neq 0} \frac{x^\intercal Ax}{x^\intercal x},$$

where S is a subspace of $\mathbb{R}^n$.

***<u>Solution</u>*:** Let us define the *Rayleigh quotient* of a non-zero vector x by

$$\mathcal{R}_A(x) = \frac{x^\intercal Ax}{x^\intercal x}.$$

Let S be an i-dimensional subspace. By a dimensionality argument, the intersection of S with span $\{f_i, \ldots, f_n\}$ is nontrivial. Let x be a non-zero vector in this intesection and write

$$x = \sum_{j=i}^{n} a_j f_j.$$

We have that

$$\mathcal{R}_A(x) = \frac{x^\intercal Ax}{x^\intercal x} = \frac{\sum_{j=i}^{n} \lambda_j a_j^2}{\sum_{j=i}^{n} a_j^2} \geq \lambda_i.$$

Now, we see that

$$\max_{x\in S, x\neq 0} \frac{x^\intercal Ax}{x^\intercal x} = \max_{x\in S, x\neq 0} \mathcal{R}_A(x) \geq \lambda_i.$$

As S was arbitrarily chosen, we have that

$$\min_{S:\dim S=i} \quad \max_{x\in S, x\neq 0} \mathcal{R}_A(x) \geq \lambda_i.$$

The equality follows by choosing $S = \text{span}\{f_1, \ldots, f_i\}$ and $x = f_i$.

The second statement is proven is a similar fashion. Choose any $(n-i+1)$-dimensional subspace S and consider its intersection with span$\{f_1, \ldots, f_i\}$. Let $x = \sum_{j=1}^{i} a_j f_j$ be a non-zero vector in this intersection. We have the following

$$\mathcal{R}_A(x) = \frac{x^\intercal Ax}{x^\intercal x} = \frac{\sum_{j=1}^{i} \lambda_j a_j^2}{\sum_{j=1}^{i} a_j^2} \leq \lambda_i,$$

and hence

$$\min_{x\in S, x\neq 0} \frac{x^\intercal Ax}{x^\intercal x} = \min_{x\in S, x\neq 0} \mathcal{R}_A(x) \leq \lambda_i.$$

As S was arbitrarily chosen, we have that

$$\max_{S:\dim S=n-i+1} \quad \min_{x\in S, x\neq 0} \mathcal{R}_A(x) \leq \lambda_i.$$

The equality follows by choosing $S = \text{span}\{f_i, \ldots, f_n\}$ and $x = f_i$.

Problem 2 (challenging): We know that every rule for clustering must display some strange behaviour. In this problem, you will prove this for partitioning a weighted graph $G = (V, E, w)$ by minimising the conductance h_G. In particular, you will consider dividing a graph into two pieces by finding the set $S \subseteq V$ with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ minimising

$$h_G(S) \triangleq \frac{w(S, V \setminus S)}{\mathrm{vol}(S)}.$$

You need to show that it is possible to split a cluster by adding an edge to the cluster or by increasing the weight of an edge inside the cluster. That is, construct a graph G so that if S is the set minimising h_G, there is an edge you can add between the vertices of S, or an edge between the vertices of S whose weight you can increase, so that after you do this the set S' minimising h_G is a proper subset of S. This goal will consists of the following two tasks:

1. Describe your graph G, the set S minimising h_G, and prove your claim.
2. Describe the edge you add or whose weight you increase to produce a new graph G'; describe the set S' minimising $h_{G'}$ in the modified graph and prove your claim.

***Solution*:**

1. Let G be a path on 6 vertices labeled from 1 to 6 such that every vertex (except for the endpoints) i is connected to vertices $i-1$ and $i+1$. Also consider the edge weights to be $50, 10, 100, 15, 10000$, for the respective edges $\{1,2\}$, $\{2,3\}$, $\{3,4\}$, $\{4,5\}$, $\{5,6\}$. Let $S \subseteq V$ be a set minimising $h_G(S)$. First, we will show that we can assume without loss of generality that S is a path. Suppose S is not a path. Then we can write it as $S = S_1 \cup S_2$ such that there are no edges between S_1 and S_2. This implies that $\mathrm{vol}(S) = \mathrm{vol}(S_1) + \mathrm{vol}(S_2)$. Moreover, we see that

$$\begin{aligned} w(S, V \setminus S) &= w(S_1 \cup S_2, V \setminus S) \\ &= w(S_1, V \setminus S) + w(S_2, V \setminus S) = w(S_1, V \setminus S_1) + w(S_2, V \setminus S_2). \end{aligned}$$

   Hence,

$$\begin{aligned} h_G(S) \triangleq \frac{w(S, V \setminus S)}{\mathrm{vol}(S)} &= \frac{w(S_1, V \setminus S_1) + w(S_2, V \setminus S_2)}{\mathrm{vol}(S_1) + \mathrm{vol}(S_2)} \\ &\geq \min\{h_G(S_1), h_G(S_2)\}, \end{aligned}$$

   where the last inequality comes from the general inequality $\frac{a_1+a_2}{b_1+b_2} \geq \min\{\frac{a_1}{b_1}, \frac{a_2}{b_2}\}$, for all $a_1, a_2, b_1, b_2 \geq 0$ and some $b_i > 0$. We can now reason inductively to conclude that S is at least as good as a path.

   Second, it is easy to see that S must contain at least 2 vertices as $h_G(\{i\}) = 1$ for every vertex i, which is clearly not minimised. Moreover, S cannot contain both vertices 5 and 6 as that would make $\mathrm{vol}(S)$ too large. Similarly, if S contains vertex 5, the vaue $h_G(S)$ is very close to 1 because both the numerator and the denominator will be dominated by the weight of the edge $\{5,6\}$ since it is much larger compared to the other edge weights. This reduces the candidates for set S to 6 choices : $\{1,2\}$, $\{2,3\}$, $\{3,4\}$, $\{1,2,3\}$, $\{2,3,4\}$, $\{1,2,3,4\}$. A direct calculation shows that the minimum $h_G(S) = \frac{15}{335}$ is achieved for the last set.

2. We can increase the value of the edge $\{1,2\}$ from 50 to 500 and repeat the analysis from the previous part, then see that the minimum cut in this case is achieved for $S' = \{1,2\}$ which is a proper subset for S.