CMSC 25025 / STAT 37601

Assignment 2 Solutions

1. *PCA* (25 points)

   (a) The problem of fitting the best k-dimensional subspace to data $x_1, \ldots, x_n \in \mathbb{R}^d$ can be written as the optimization

$$\min_{\mu, \{\lambda_i\}, V_k} \sum_{i=1}^{n} \|x_i - \mu - V_k \lambda_i\|^2$$

where V_k is an $d \times k$ orthogonal matrix. Show that an optimum over the variables $\mu \in \mathbb{R}^d$ and $\lambda_i \in \mathbb{R}^k$ is given by

$$\widehat{\mu} = \overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\widehat{\lambda}_i = V_k^T (x_i - \overline{x}_n).$$

Show that $\widehat{\mu}$ is not unique, and characterize the set of possible solutions.

Set $f(\mu, \lambda_1, \cdots, \lambda_n) = \sum_{i=1}^{n} \|x_i - \mu - V_k \lambda_i\|^2$. Take derivative respect to μ, we have

$$\frac{\partial f}{\partial \mu} = 2n\mu - 2\sum_{i=1}^{n} x_i + 2V_k \sum_{i=1}^{n} \lambda_i.$$

Then we have

$$\widehat{\mu} = \bar{x} - V_k \bar{\lambda}.$$

Take derivative respect to λ_i, we have

$$\frac{\partial f}{\partial \lambda_i} = 2V_k^T V_k \lambda_i + 2V_k^T (\mu - x_i),$$

Plug $\widehat{\mu} = \bar{x} - V_k \bar{\lambda}$ in and note that $V_k^T V_k = I_k$, we have

$$\widehat{\lambda}_i = V_k^T (x_i - \bar{x}) + \bar{\lambda}.$$

For the solution above, $\bar{\lambda}$ is a free parameter and can be set as any value. Set it to be 0, then $\widehat{\mu} = \bar{x}$, $\widehat{\lambda}_i = V_k^T (x_i - \bar{x})$ is an optimum.

Obviously, $\bar{x}$ is not the unique optimum for μ, the set of possible solutions is $\{(\mu, \lambda_1, \cdots, \lambda_n) : \mu = \bar{x} - V_k \theta, \lambda_i = V_k^T (x_i - \bar{x}) + \theta, i = 1, \cdots, n, \theta \in \mathbb{R}^k\}$

(b) Let S be a non-negative $d \times d$ symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_d$. Let V be the orthogonal matrix of eigenvectors of S. Let $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$ be the diagonal matrix of the eigenvalues so that $S = V\Lambda V^t$. Show that the $d \times k$ matrix $V^{[k]}$ of the first k columns of V maximizes:

$$tr(U^t S U),$$

over all orthogonal $d \times k$ matrices.

Note that

$$tr(U^T S U) = tr(U^T V \Lambda V^T U) = tr(\Lambda V^T U U^T V).$$

Let $Q = V^T U \in R^{d \times k}$. It suffices to show that, over all orthogonal $d \times k$ matrices, the matrix that includes the first k columns of identity matrix can maximize the above quantity. In fact, let $Q = (q_1, \ldots, q_k)$ and we have

$$tr(\Lambda Q Q^T) = \sum_{i=1}^{k} q_i^T \Lambda q_i = \sum_{j=1}^{d} \lambda_j (\sum_{i=1}^{k} q_{i,j}^2) \triangleq \sum_{j=1}^{d} \lambda_j a_j.$$

Since Q has orthonormal columns, $a_j \leq 1$ and $\sum_{j=1}^{d} a_j = k$. Therefore, the above right hand side is upper bounded by $\sum_{j=1}^{k} \lambda_j$, which can be achieved by letting $Q = I_k$. This completes the proof.

2. *Ncuts and eigenvalue problems:* (25 points) We have a weighted graph with a set V of n vertices and symmetric weight matrix $W_{ij}, i, j = 1, \ldots, n$. The edge set of the graph is $E = \{(i,j) : W_{ij} > 0\}$. with non-negative entries. Define the degree of a vertex as $d_i = \sum_j W_{ij}$. The degree of a subset A of vertices is the sum of their degrees: $d(A) = \sum_{i \in A} d_i$, and $d = (d_1, \ldots, d_n)$. Let A, B be a partition of the n nodes and define the normalized cut

$$\text{Ncut}(A, B) = \text{cut}(A, B) \left(\frac{1}{d(A)} + \frac{1}{d(B)} \right),$$

where $\text{cut}(A, B) = \sum_{i \in A, j \in B} W_{ij}$.

(a) Define the Laplacian of the graph as $D - W$, where $D = \text{diag}(d)$. Show that $f^t L f = \frac{1}{2} \sum_{ij} W_{ij} (f_i - f_j)^2$. And explain why the constant vector of 1's, $\mathbf{1}$ is the eigenvector of L with eigenvalue 0.

$f^t (D - W) f = \sum_i f_i^2 d_i - \sum_{ij} f_i W_{ij} f_j$, and

$$\begin{aligned} 1/2 \sum_{ij} W_{ij} (f_i - f_j)^2 =& 1/2 \sum_{ij} W_{ij} f_i^2 + 1/2 \sum_{ij} W_{ij} f_i^2 - \sum_{ij} W_{ij} f_i f_j \\ =& \sum_i f_i^2 d_i - \sum_{ij} W_{ij} f_i f_j. \end{aligned}$$

The right hand side is always non-negative so that L is non-negative definite. A simple computation yields $D\mathbf{1} = W\mathbf{1}$.

(b) Assume the graph is connected, i.e for any two vertices i, j there is a path in the graph between i and j. Show that any eigenvector with eigenvalue 0 has to be of the form $c\mathbf{1}$ for some $c \in R$.

If f is not constant, there are two indices i, j s.t. $f_i \neq f_j$. Let $k_1, \ldots, k_l$ be a path in the graph between i and j, with $k_1 = i, k_l = j$. Since the graph is connected there must be a consecutive pair k_r, k_{r+1} such that $f_{k_r} \neq f_{k_{r+1}}$ otherwise $f_i = f_{k_1} = f_{k_l} = f_j$. The pair k_r, k_{r+1} contributes a strictly positive term to the sum $W_{k_r,k_{r+1}}(f_{k_r} - f_{k_{r+1}})^2$.

(c) Let $f \in R^n$ have only two values, i.e. $f_i \in \{a, b\}, a \neq b$. Let $A = \{i : f_i = a\}, B = A^c$. Assuming $f^t d = 0$, what is the unique value of a/b.

Since $f^t d = 0$ we have $0 = \sum_{i\in A} ad_i + \sum_{i\in B} bd_i = a \cdot d(A) + b \cdot d(B)$, implying that $a/b = -d(B)/d(A)$. Let's assume that $a = 1/d(A), b = -1/d(B)$

(d) Let f be as in 2c, show that

$$\frac{f^t L f}{f^t D f} = Ncut(A, B).$$

And conclude that finding the minimum Ncut is equivalent to minimizing

$$\frac{f^t L f}{f^t D f}, \text{ s.t. } f^t D\mathbf{1} = 0, \text{ and } f_i \in \{a, b\}.$$

$f^t L f = 1/2 \sum_{ij} W_{ij}(f_i - f_j)^2 = \sum_{i\in A, j\in B} W_{ij}(1/d(A) + 1/d(B))^2$, because the terms in the sum with indices both in A or both in B are 0. Then the denominator is $f^t D f = \sum_{i\in A} d_i/d(A)^2 + \sum_{i\in B} d_i/d(B)^2 = d(A)/d(A)^2 + d(B)/d(B)^2 = (1/d(A) + 1/d(B))$. So the ratio is equal to Ncut(A, B).

(e) Show that if f_* minimizes the relaxed problem (i.e. dropping the last constraint) if and only if $f_* = D^{-1/2}u_*$ where u_* minizes $\frac{u^t \widetilde{L} u}{u^t u}$ subject to $u^t D^{1/2}\mathbf{1} = 0$, where $\widetilde{L} = I - D^{-1/2} W D^{-1/2}$.

For any f, if we take $u = D^{1/2} f$ and plug it in the ratio we get:

$$\frac{u^t D^{-1/2}(D - W)D^{-1/2} u}{u^t D^{-1/2} D D^{-1/2} u} = \frac{u^t \widetilde{L} u}{u^t u},$$

and $u^t D^{-1/2} D\mathbf{1} = u^t D^{1/2}\mathbf{1}$. So the relaxed minimization problem is equivalent to minimizing

$$\frac{u^t \widetilde{L} u}{u^t u}, \text{s.t..} u^t D^{1/2}\mathbf{1} = 0.$$

(f) Show that $D^{1/2}\mathbf{1}$ is the eigenvector of $\widetilde{L}$ with eigenvalue 0, and u^* is the eigenvector of $\widetilde{L}$ with second smallest eigenvalue.

Simple computation shows that since $\mathbf{1}$ is the unique 0 eigenvector of L, $D^{1/2}\mathbf{1}$ is the unique eigenvalue of $\widetilde{L}$ with value 0. The miminizer u_* is the eigenvector of $\widetilde{L}$. This can be seen based on problem 2. Take $k = 2$ and let $u_0 = D^{1/2}\mathbf{1}/|D^{1/2}\mathbf{1}$ and $u_{*,1} = u_*/|u_*|$. Then $\widetilde{L}u_0 = 0$,, and and $u_{1,*}$ minimizes u^tLu over all unit vectors orthogonal to u_0. By definition $U = [u_0, u_{*,1}]$ is an orthogonal matrix and

$$U^t\widetilde{L}U = u_0^t\widetilde{L}u_0 + u_{*,1}\widetilde{L}u_{*,1} = 0 + u_{*,1}\widetilde{L}u_{*,1}.$$

So clearly U minimizes U^tLU over all two column orthogonal matrices, so the two columns of U are the bottom two eigenvectors of $\widetilde{L}$.