STA 37601 HW2

Albert Chu

April 8, 2025

1. (a) Begin by considering a given i with the term

$$|x_i - \mu - V_k\lambda_i|^2$$

Since V_k has orthonormal columns, λ_i can be optimized for a fixed μ by projecting $x_i - \mu$ onto the subspace of the columns of V_k. In particular,

$$\hat{\lambda}_i = V_k^T(x_i - \mu)$$

Substituting this in yields

$$x_i - \mu - V_k V_k^T(x_i - \mu) = (I - V_k V_k^T)(x_i - \mu) = (I - V_k V_k^T)x_i - (I - V_k V_k^T)\mu$$

This shows us that the best μ will minimize the distance between the projection of the data, $(I - V_k V_k^T)x_i$, and $(I - V_k V_k^T)\mu$. We know that in this least squares problem, the minimizer is then given by

$$\hat{\mu} = \frac{1}{n}\sum x_i \rightarrow \hat{\lambda}_i = V_k^T(x_i - \bar{x})$$

To see why $\hat{\mu}$ is not unique, consider for any $c \in R^k$,

$$\mu' = \hat{x} + V_k c$$

$$\lambda_i' = V_k^T(x_i - \hat{x}) - c$$

And observe that

$$\mu' + V_k\lambda_i' = (\bar{x} + V_k c) + V_k(V_k^T(x_i - \bar{x}) - c) = \bar{x} + V_k V_k^T(x_i - \bar{x})$$

In essence, the shift of c cancel sout and $\mu' + V_k\lambda_i'$ is equivalent here to $\hat{\mu} + V_k\hat{\lambda}_i$. By this logic, we can describe the entire set of possible solutions as that obtained by adding any vector from the column space of V_k to μ and subtracting the corresponding vector from each λ_i to get the same results:

$$\{(\mu, \lambda_i) | \mu = \bar{x} + V_k c, \lambda_i = V_k^T(x_i - \bar{x}) - c, \forall c \in R^k\}$$

(b) We want to maximize $tr(V_k^t C V_k)$ over all orthonormal V_k. We can substitute the spectral decomposition,

$$tr(V_k^T C V_k) = tr(V_k^T U \Lambda U^T V_k) = tr((V_k U)^T \Lambda (U^T V_k))$$

To get the maximum trace, we want to get the biggest coefficients for the largest eigenvalues λ_i, which is achieved with coefficients of 1 for the first k eigenvalues. In particular, we want the above to equal $\sum_{i=1}^k \lambda_i$. Notice that we can achieve this by choosing $V_k = U_k$ where U_k is the matrix of the first k columns of U since $U^T U_k = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$

2. (a) Let $L = D - W$. For a given vector f, we then have

$$f^T L f = f^T (D - W) f = f^T D f - f^T W f$$

Since D is a diagonal matrix of d, we can rewrite the first term

$$f^T D f = \sum_i^n d_i f_i^2 = \sum_i^n \left(\sum_j W_{ij} \right) f_i^2 = \sum_{i,j} W_{ij} f_i^2$$

Similarly, the second term becomes

$$f^T W f = \sum_{i,j} W_{ij} f_i f_j$$

Since W is symmetric,

$$f^T L f = \sum_{i,j} W_{ij} f_i^2 - \sum_{i,j} W_{i,j} f_i f_j = \frac{1}{2} \sum_{i,j} W_{ij} (f_i - f_j)^2$$

Now, consider the constant vector of 1s. Notice that for any pair of nodes i and j,

$$(1_i - 1_j)^2 = (1 - 1)^2 = 0$$

So,

$$1^T L 1 = \frac{1}{2} \sum_{ij} W_{ij} (1 - 1)^2 = 0$$

Since L is symmetric and positive semidefinite, this implies that $L1 = 0$.

(b) Let f be an eigenvector of 0, i.e. $Lf = 0$. This implies from the above that $f^T L f = 0$. Since every term in $\frac{1}{2} \sum_{i,j} W_{ij} (f_i - f_j)^2$ is non-negative, the sum can only be zero if each term is 0. This implies that for every pair of vertices i, j, we must have $(f_i - f_j)^2 = 0 \implies f_i = f_j$.

Given the assumption that there is a path between any i and j, we can apply the above result repeatedly along the path to conclude that $f_i = f_j$ for all i, j. This therefore implies that f must be a constant vector. In particular, $\exists c \in R$ st $f = c1$.

(c) Denote the degree of vertex i as d_i and $d(A), d(B)$. We are to assume that $f^t d = \sum_i^n d_i f_i = 0$. We can split the sum,

$$\sum_{i\in A} d_i a + \sum_{i\in B} d_i b = ad(A) + bd(B) = 0$$

We can then solve this for our ratio if the graph and a, b are nontrivial,

$$\frac{a}{b} = -\frac{d(B)}{d(A)}$$

(d) From before, we have that $f^T L f = \frac{1}{2}\sum_{i,j} W_{ij}(f_i - f_j)^2$ and we are to assume that f takes values either a or b. Since f_i is constant within A or B, $(f_i - f_j)^2 = 0$ within those, meaning only edges between A and B contribute anything. In particular, $(f_i - f_j)^2 = (a - b)^2$,

$$f^T L f = \frac{1}{2}\sum_{i\in A, j\in B} W_{ij}(a-b)^2 + \frac{1}{2}\sum_{i\in B, j\in A} W_{ij}(a-b)^2$$

Since W is symmetric, this simplifies to

$$f^T L f = \sum_{i\in A, j\in B} W_{ij}(a-b)^2 = cut(A, B)(a-b)^2$$

Next, we consider

$$f^T D f = \sum_i^n d_i f_i^2 = \sum_{i\in A} d_i a^2 + \sum_{i\in B} d_i b^2 = a^2 d(A) + b^2 d(B)$$

We showed above that assuming $f^T d = 0$ implies $\frac{a}{b} = -\frac{d(B)}{d(A)}$ Then,

$$a - b = -\frac{d(B)}{d(A)} b - b = -b\left(\frac{d(B) - D(A)}{d(A)}\right)$$

$$(a-b)^2 = b^2 \frac{(d(A)^2 + d(B)^2)}{d(A)^2}$$

We also see that

$$a^2 d(A) + b^2 d(B) = b^2\left(\frac{d(B)^2}{d(A)^2} d(A) + d(B)\right) = b^2 \frac{d(B)(d(B) + d(A))}{d(A)}$$

We can now write our final formula as

$$\frac{f^T L f}{f^T D f} = \frac{(a-b)^2 cut(A,B)}{a^2 d(A) + b^2 d(B)} = \frac{d(A) + d(B)}{d(A)} \cdot \frac{cut(A,B)}{d(B)} = cut(A,B)\left(\frac{1}{d(A)} + \frac{1}{d(B)}\right)$$

This allows us to conclude that minimizing the ncut is equivalent to finding f that satisfies $f^T D 1 = 0$ and minimizes $\frac{f^T L f}{f^T D f}$.

(e) Define $f = D^{-1/2}u$ and notice that

$$f^T Df = (D^{-1/2}u)^T D(D^{-1/2}u) = u^T u$$

and since $L = D - W$,

$$f^T Lf = (D^{-1/2}u)^T L(D^{-1/2}u) = u^T(D^{-1/2}LD^{-1/2})u = u^T(D^{-1/2}(D-W)D^{-1/2})u = u^T\tilde{L}u$$

Putting these together, we have that

$$\frac{f^T Lf}{f^T Df} = \frac{u^T\tilde{L}u}{u^T u}$$

Our original constraint of $f^T D1 = 0$ can be transformed as well,

$$(D^{-1/2}u)^T D1 = u^T D^{1/2}1 = 0$$

(f) Let $v = D^{1/2}1$, then

$$\tilde{L}v = (I - D^{-1/2}WD^{-1/2})(D^{1/2}1) = D^{1/2}1 - D^{-1/2}W1$$

Notice that $W1 = d$, so $D^{-1/2}W1 = D^{-1/2}d$ and $D^{-1/2}d = D^{1/2}$ by definition, giving us

$$\tilde{L}(D^{1/2}1) = D^{1/2}1 - D^{1/2}1 = 0$$

and so $D^{1/2}1$ is the eigenvector of $\tilde{L}$ with eigenvalue 0.
Since $D^{1/2}1$ is the eigenvector corresponding to the smallest eigenvalue, the orthogonality constraint $u^T D^{1/2}1 = 0$ forces u to be orthogonal to this trivial solution. The smallest eigenvalue available for this is λ_2 and to minimize our ratio, we want to put all the 'weight' on the smallest eigenvalue, i.e. $u* = u_2$ with $\tilde{L}u_2 = \lambda_2 u_2$, which gives

$$\frac{u*^T\,\tilde{L}u*}{u*^T\,u*} = \lambda_2$$

3. See coding submission.