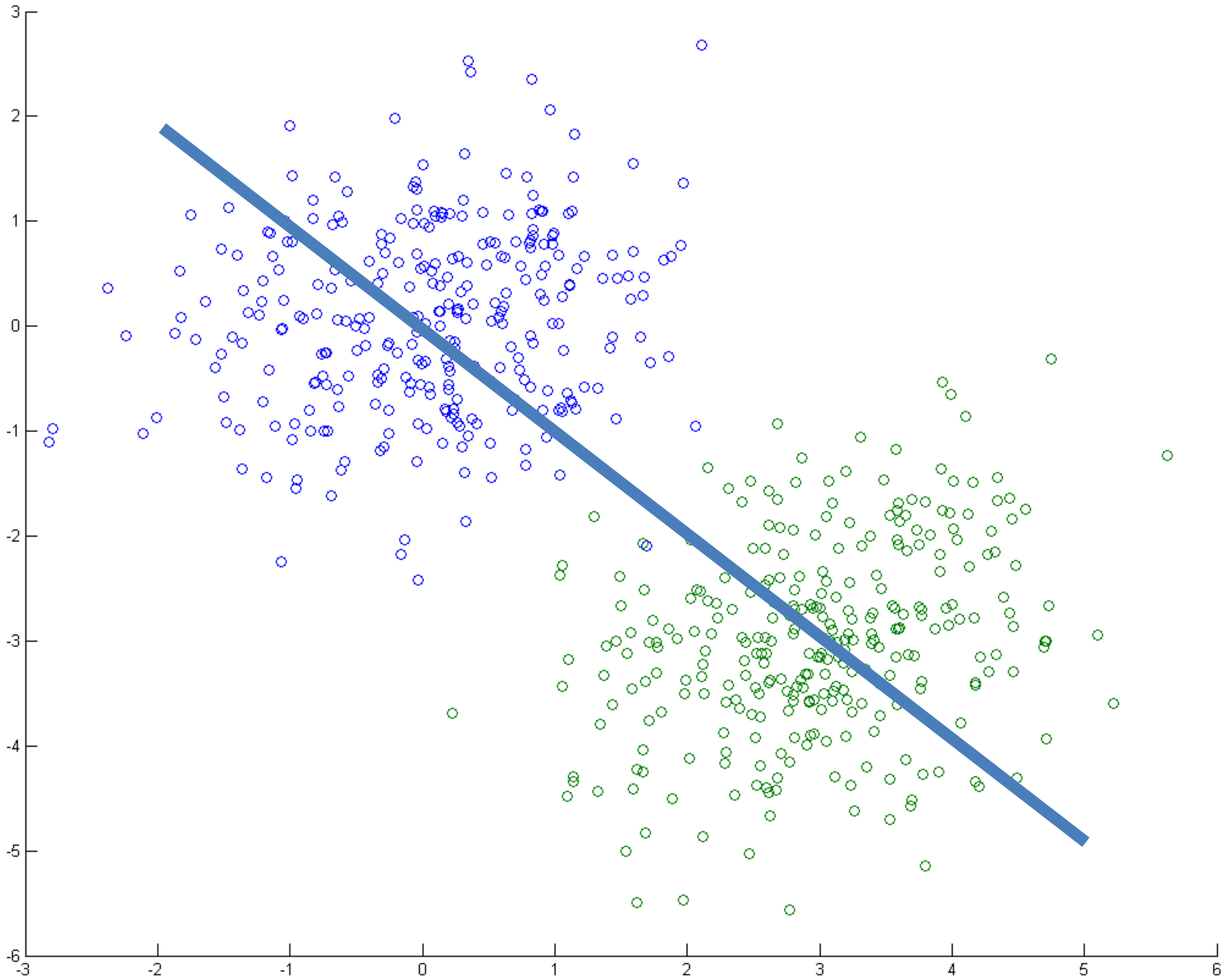
Machine Learning – Assignment 5

103062528 林玉山

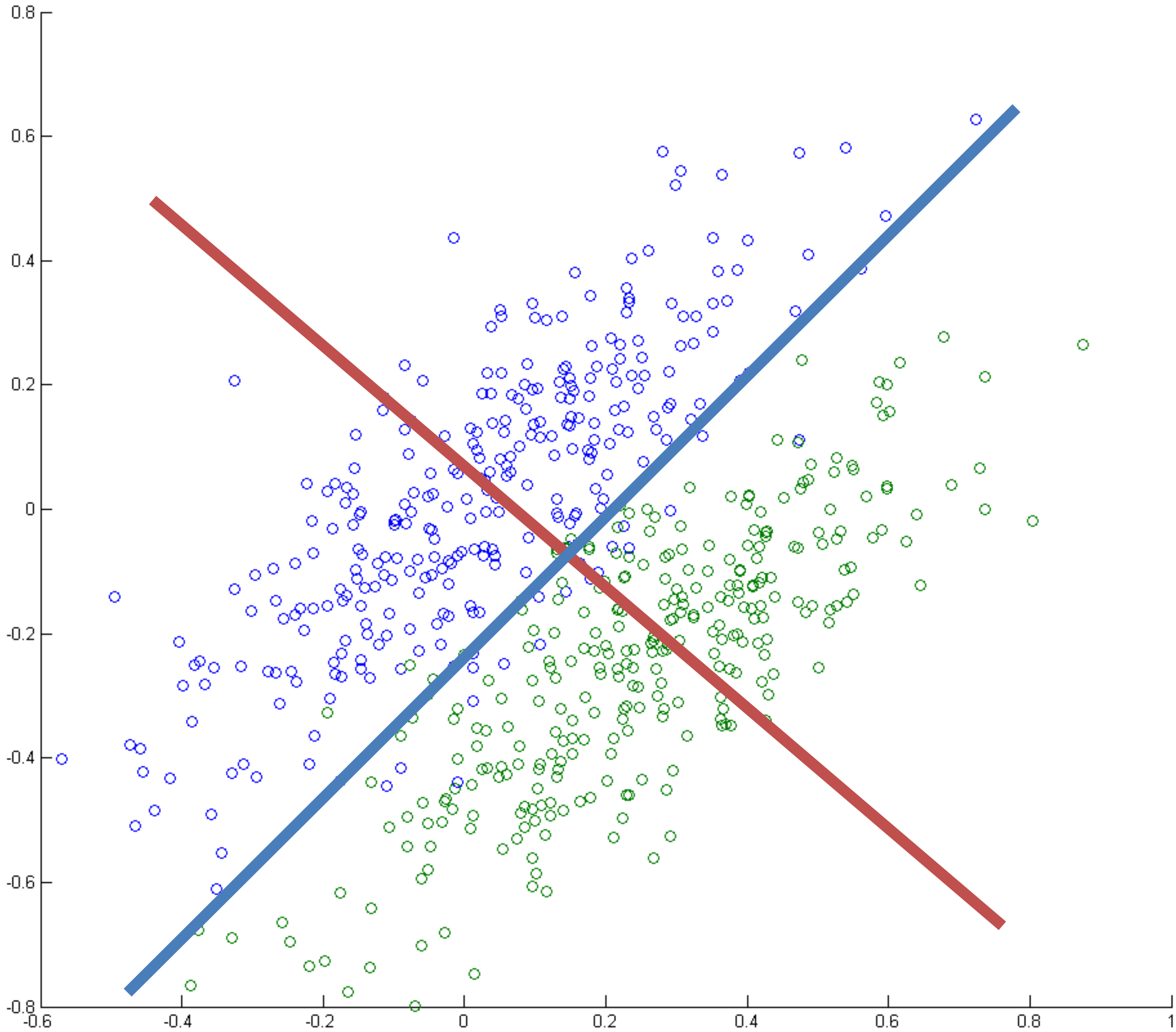
1. 假設 d=2，畫出兩組 Normal Distribution 的 Data Set 以符合下述兩種狀況

(a) 做 PCA 得到的 vector，與 LDA 得到的 vector 具有相同方向



PCA 會盡可能地將使資料點投影到新的 vector 上時能夠區別開來，因此會選擇上圖軸藍軸的方向。LDA 則會盡可能地將「不同群」的資料點分開，因此也會趨向於藍軸。最後兩種方式都會獲得藍軸這個 vector。

(b) 做 PCA 得到的 vector，與 LDA 得到的 vector 互相正交



PCA 為了盡可能使資料點分開，而這份資料投射到藍軸後比較容易分別，因此會採用藍軸。不過 LDA 為了能夠使不同群資料分開，也就是盡可能地讓藍色與綠色在新軸上也能夠區別，所以會趨向於使用紅軸。

若將上述資料視為橢圓，PCA 會趨向於取長軸。如果資料分群的界線剛好是短軸的話，那就會產生 PCA 與 LDA 選的軸正交的狀況。

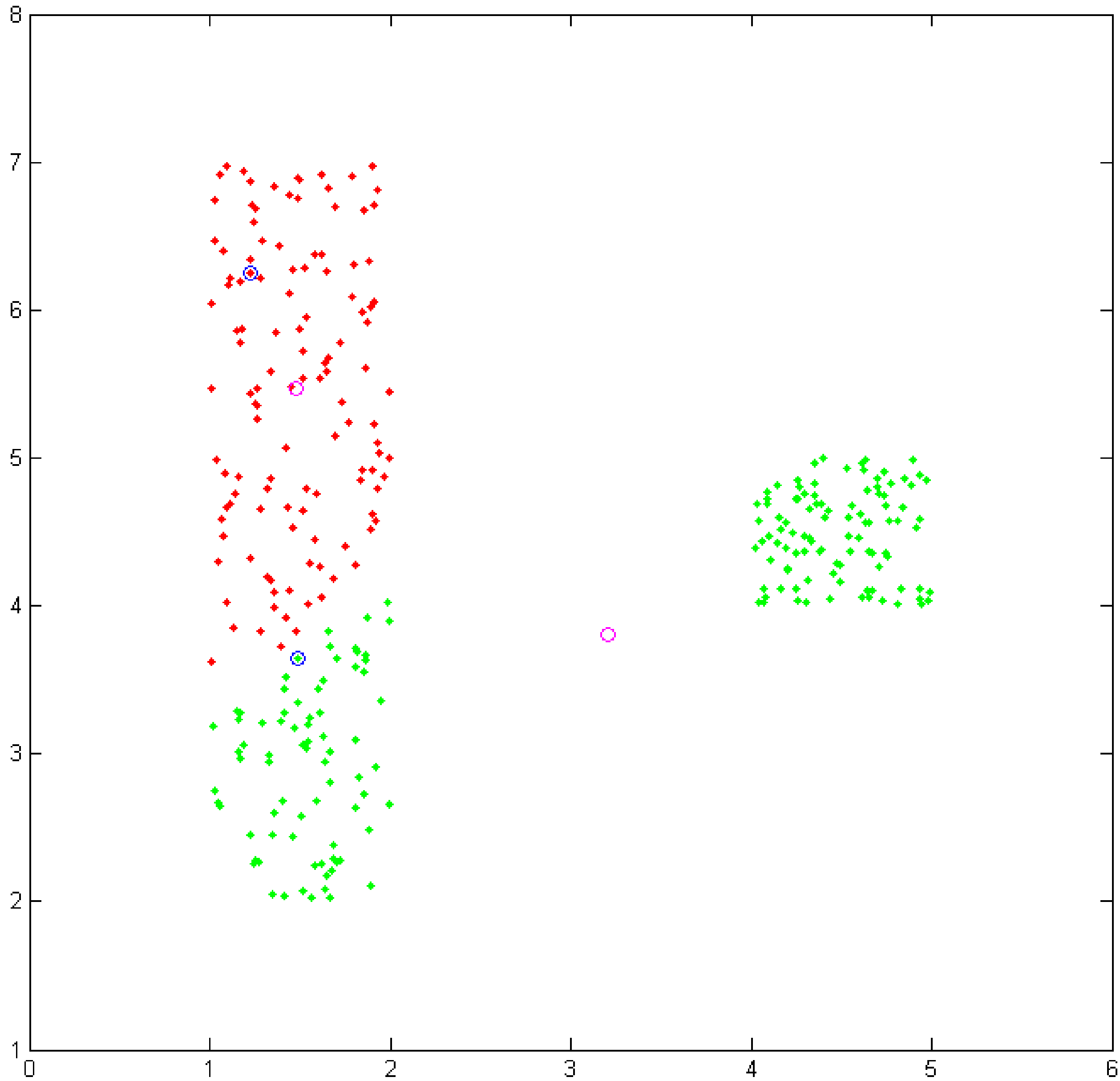
2. 延伸現有的 Kernel PCA，使之可以不用 center

(不會寫，來不及複習 QQ)

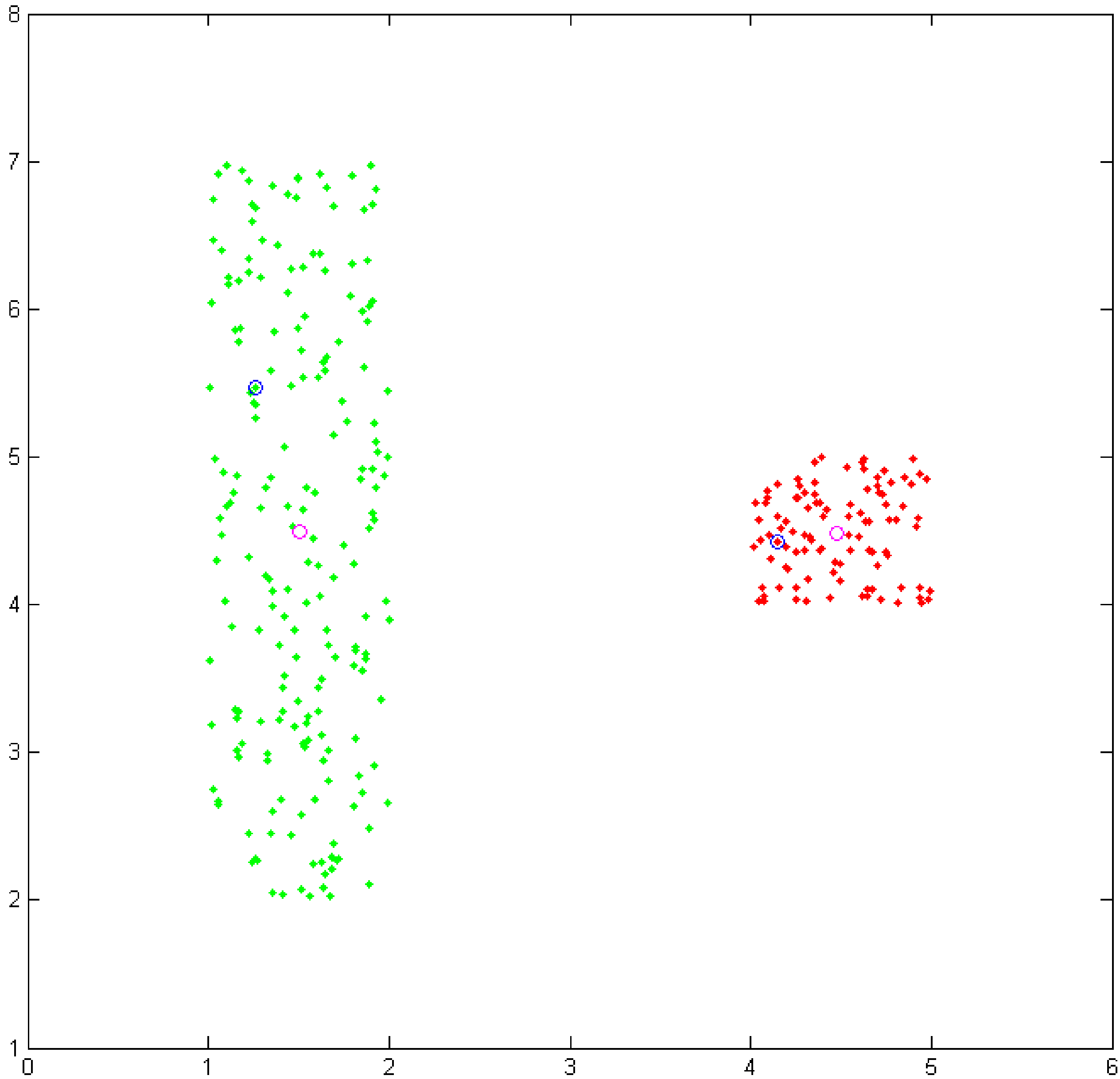
3. 令 d=2, k=2，畫出一張一開始的 mean 選得不好的圖，以及一張選得好的圖

下兩張圖為同一組 data set，分為紅色與綠色兩群。藍色圈圈為一開始的 means，紫色為最後收斂時的 means。

第一張圖為一開始的兩個 means 都從左側那群點中取出。從最後收斂的結果可以看出來，左半部的下半段被分進了右邊那群中。很明顯就不是我們要的結果。



第二張圖則一開始就從左右兩群個挑一個點出來做為初始值。可以看見這樣的情況下，左右兩群就被正確地分開了。並且最後得到的 means，也正好在兩群的正中央。正是我們期望的結果。由此可以得到「一開始挑選的點會大大影響結果」這個結論。



4. 為什麼在 Hierarchical Clustering 中，我們可以捨去部分的 subtree？

在 Hierarchical Clustering 中，當我們選擇了一個 distance，h 作為斷點，切出了數個 clusters 之後。那些數量極少的 clusters，之所以沒有被併進其他 clusters 中，是因為他們距離那幾個 clusters 都有至少 h 以上的距離。因此就算我們把這些數量極少的 clusters 都去除掉，也不會影響其他 clusters 的結果，因為他們的距離都太遠，不足以影響到較大的 clusters。

5. 令有一 Graph Laplacian Matrix L，請證明下列性質

(a) 證明對於任意向量 $\boldsymbol{f \epsilon R^N}$，$\boldsymbol{f^T L f = \frac{1}{2} \Sigma_{i,j=1}^N s_{i,j} (f_i - f_i)^2}$

首先，我們先計算 L

$$L = D - S = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_N \end{bmatrix} - \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \cdots & s_{N,N} \end{bmatrix}$$

$$= \begin{bmatrix} d_1 - s_{1,1} & -s_{1,2} & \cdots & -s_{1,N} \\ -s_{2,1} & d_2 - s_{2,2} & \cdots & -s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -s_{N,1} & -s_{N,2} & \cdots & d_N - s_{N,N} \end{bmatrix}$$

接著，我們展開 $f^T L f$

$$f^T L f = \boxed{\begin{bmatrix} f_1 & f_2 & \cdots & f_N \end{bmatrix}} \begin{bmatrix} d_1 - s_{1,1} & -s_{1,2} & \cdots & -s_{1,N} \\ -s_{2,1} & d_2 - s_{2,2} & \cdots & -s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -s_{N,1} & -s_{N,2} & \cdots & d_N - s_{N,N} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}$$

當我們計算綠色部分乘上 L 時，我們令每一個 column 的結果為 p_i

$$p_i = f_i d_i - \sum_{t=1}^{N} f_t s_{t,i}$$

此時 $f^T L f$ 變為：

$$f^T L f = \begin{bmatrix} p_1 & p_2 & \cdots & p_N \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} = f_1 p_1 + f_2 p_2 \cdots + f_N p_N$$

我們再將 p_i, d_i 各自代入，其中對於每一個 $f_i p_i$

$$f_i p_i = f_i \left(f_i d_i - \sum_{t=1}^{N} f_t s_{t,i} \right) = f_i \left(f_i \sum_{t=1}^{N} s_{i,t} - \sum_{t=1}^{N} f_t s_{t,i} \right) = f_i^2 \sum_{t=1}^{N} s_{i,t} - f_i \sum_{t=1}^{N} f_t s_{t,i}$$

因此 f^TLf 為：

$$f^TLf = \sum_{i=1}^{N} f_i^2 \sum_{t=1}^{N} s_{i,t} - f_i \sum_{t=1}^{N} f_t s_{t,i}$$

$$= \sum_{i=1}^{N} \sum_{t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i}$$

$$= \sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i}$$

$$= \sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i}$$

$$= \frac{1}{2} \times 2 \sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i}$$

$$= \frac{1}{2} (\sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i} + \sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i})$$

在上式中，我們讓左式(綠色)與右式之中，index 相反的項相加，可以得到下式

$$= \frac{1}{2} \sum_{i,t=1}^{N} f_i^2 s_{i,t} - f_i f_t s_{t,i} + f_t^2 s_{t,i} - f_t f_i s_{i,t}$$

又因為矩陣 S 是 Symmetric 的，所以對於 $\mathrm{i}, \mathrm{t} = 1 \sim \mathrm{N},\ s_{t,i} = s_{i,t}$

$$= \frac{1}{2} \sum_{i,t=1}^{N} s_{i,t} (f_i^2 - f_i f_t + f_t^2 - f_t f_i)$$

$$= \frac{1}{2} \sum_{i,t=1}^{N} s_{i,t} (f_i^2 - 2 f_i f_t + f_t^2)$$

$$= \frac{1}{2} \sum_{i,t=1}^{N} s_{i,t} (f_i - f_t)^2$$

將 t 換成 j 即為原式，故得證。

(b) 試證明 $\mathbf{tr}(\boldsymbol{f}^{\top}\boldsymbol{L}\boldsymbol{f}) = \mathbf{2}\boldsymbol{RatioCut}(\boldsymbol{G}, \overline{\boldsymbol{G}})$

目前已知 (注意 $f^{\top}Lf$ 是純量)：

$$\mathrm{tr}(f^{\top}Lf) = \mathrm{tr}\left(\frac{1}{2}\sum_{i,j=1}^{N} s_{i,j}(f_i - f_j)^2\right) = \frac{1}{2}\sum_{i,j=1}^{N} s_{i,j}(f_i - f_j)^2$$

接著對於第 i 與第 j 個 data point，一共會有四種情形：

(1) i 與 j 同屬於 G

(2) i 屬於 G，j 不屬於 G

(3) i 不屬於 G，j 屬於 G

(4) i 與 j 皆不屬於 G

因此我們可以依照四種將原式中的 Σ 拆開：

$$= \frac{1}{2}\Big(\sum_{i\in G, j\in G} s_{i,j}(f_i - f_j)^2 + \sum_{i\in G, j\in \bar{G}} s_{i,j}(f_i - f_j)^2 + \sum_{i\in \bar{G}, j\in G} s_{i,j}(f_i - f_j)^2 + \sum_{i\in \bar{G}, j\in \bar{G}} s_{i,j}(f_i - f_j)^2\Big)$$

此時代入下式

$$f_i = \begin{cases} \sqrt{\dfrac{|\bar{G}|}{N|G|}}, \mathrm{if}\ x^{(i)} \in G \\ -\sqrt{\dfrac{|G|}{N|\bar{G}|}}, \mathrm{otherwise} \end{cases}$$

$$= \frac{1}{2}\Big(\sum_{i\in G, j\in G} s_{i,j}\left(\sqrt{\frac{|\bar{G}|}{N|G|}} - \sqrt{\frac{|\bar{G}|}{N|G|}}\right)^2 + \sum_{i\in G, j\in \bar{G}} s_{i,j}\left(\sqrt{\frac{|\bar{G}|}{N|G|}} + \sqrt{\frac{|G|}{N|\bar{G}|}}\right)^2$$

$$+ \sum_{i\in \bar{G}, j\in G} s_{i,j}\left(-\sqrt{\frac{|G|}{N|\bar{G}|}} - \sqrt{\frac{|\bar{G}|}{N|G|}}\right)^2 + \sum_{i\in \bar{G}, j\in \bar{G}} s_{i,j}\left(-\sqrt{\frac{|G|}{N|\bar{G}|}} + \sqrt{\frac{|G|}{N|\bar{G}|}}\right)^2\Big)$$

$$=\frac{1}{2}(\sum_{i\in G,j\in\bar{G}} s_{i,j}\left(\sqrt{\frac{|\bar{G}|}{N|G|}}+\sqrt{\frac{|G|}{N|\bar{G}|}}\right)^2+\sum_{i\in\bar{G},j\in G} s_{i,j}\left(-\sqrt{\frac{|G|}{N|\bar{G}|}}-\sqrt{\frac{|\bar{G}|}{N|G|}}\right)^2)$$

因為 $Cut(G)=\sum_{i\in G,j\in\bar{G}} s_{i,j}$，因此

$$\mathrm{tr}(f^\top Lf)=\frac{1}{2}\left(Cut(G)\left(\sqrt{\frac{|\bar{G}|}{N|G|}}+\sqrt{\frac{|G|}{N|\bar{G}|}}\right)^2+Cut(G)\left(-\sqrt{\frac{|G|}{N|\bar{G}|}}-\sqrt{\frac{|\bar{G}|}{N|G|}}\right)^2\right)$$

$$=\frac{1}{2}\left(Cut(G)\left(\frac{|\bar{G}|}{N|G|}+2\sqrt{\frac{|\bar{G}||G|}{N^2|G||\bar{G}|}}+\frac{|G|}{N|\bar{G}|}\right)+Cut(G)\left(\frac{|\bar{G}|}{N|G|}+2\sqrt{\frac{|\bar{G}||G|}{N^2|G||\bar{G}|}}+\frac{|G|}{N|\bar{G}|}\right)\right)$$

$$=Cut(G)\left(\frac{|\bar{G}|}{N|G|}+\frac{2}{N}+\frac{|G|}{N|\bar{G}|}\right)$$

$$=\frac{Cut(G)}{N}\left(\frac{|\bar{G}|}{|G|}+2+\frac{|G|}{|\bar{G}|}\right)$$

$$=\frac{Cut(G)}{N}\left(\frac{|\bar{G}|+|G|}{|G|}+\frac{|G|+|\bar{G}|}{|\bar{G}|}\right)$$

$$=\frac{Cut(G)}{N}\left(\frac{\mathrm{N}}{|G|}+\frac{\mathrm{N}}{|\bar{G}|}\right)$$

$$=\frac{Cut(G)}{|G|}+\frac{Cut(G)}{|\bar{G}|}$$

$$=2\times\frac{1}{2}\left(\frac{Cut(G)}{|G|}+\frac{Cut(G)}{|\bar{G}|}\right)$$

$$=2\times\frac{1}{2}\left(\frac{Cut(G)}{|G|}+\frac{Cut(\bar{G})}{|\bar{G}|}\right)$$

$$=2RatioCut(G,\bar{G})$$

得證。