Jan 30, 2017

Recall the two-sample statistic:

$$\mathcal{E}(A,B) = \frac{n_1 n_2}{n_1 + n_2} \left\{ \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_\alpha(a_i, b_j) - \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} D_\alpha(a_i, a_j) - \frac{1}{n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} D_\alpha(b_i, b_j) \right\}$$

$$g_\alpha(A,B) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_\alpha(a_i, b_j)$$

We saw that $g_\alpha(A,B) = \left(\frac{1}{n_1 n_2}\right) \vec{1}_{n_1}^T D_{AB} \vec{1}_{n_2}$

$$\mathcal{E}(A,B) = \frac{n_1 n_2}{n_1 + n_2} \left\{ 2\, g_\alpha(A,B) - g_\alpha(A,A) - g_\alpha(B,B) \right\}$$

Forgetting about n_1, n_2 for now:

$$2\, \vec{1}_{n_1}^T D_{AB} \vec{1}_{n_2} - \vec{1}_{n_1}^T D_{AA} \vec{1}_{n_1} - \vec{1}_{n_2}^T D_{BB} \vec{1}_{n_2}$$

which is the same as

$$\vec{1}_{(n_1+n_2)}^T \left(\begin{array}{c|c} -D_{AA} & D_{AB} \\ \hline B_{BA} & -D_{BB} \end{array} \right) \vec{1}_{(n_1+n_2)}$$

Now the - signs can be included in the vector on the left. This give a distance data matrix

$$D$$

without knowing the structure, we must solve an optimization problem of the form

$$\boxed{x^T D \vec{1}}$$

where $x_i \in \{+1, -1\}$, $x \in \mathbb{R}^{n_1+n_2}$, $\vec{1} \in \mathbb{R}^{n_1+n_2}$.

$\mathcal{E}$ is finite if $A \overset{D}{=} B$.

$\mathcal{E} \to \infty$ if $A \overset{D}{\neq} B$.

Thus I guess we have

$$\boxed{\begin{aligned} &\min x^T D \vec{1} \\ &\text{s.t. } x_i^2 = 1 \end{aligned}}$$