each source domain, which are must-link constraints and cannot-link constraints. There is no assumption posed on the pairwise constraints. In general, the pairwise constraints are randomly selected by the users. In the $s$-th source domain, the set of must-link constraints is represented as $M^s = \{m^s_1, m^s_2, \ldots, m^s_{r_s}\}$, and the set of cannot-link constraints is denoted by $C^s = \{c^s_1, c^s_2, \ldots, c^s_{l_s}\}$, where $r_s$ and $l_s$ denote the number of must-link constraints and cannot-link constraints, respectively. More precisely, $m^s_i$ consists of a pair of points belonging to the same class while $c^s_i$ consists of a pair of points belonging to different classes. For the target domain, we assume that all the data points are unlabeled. Moreover, we assume that each domain has $K$ clusters, and the $K$ clusters in a domain conceptually correspond to the $K$ clusters in another domain. The output of our method is a cluster indicator matrix $Q_T$ for the target domain and a $d \times l$ transformation matrix $W$ with an orthogonal constraint $W^TW = I$. $W$ consists of $l$ projective vectors $\{w_1, w_2, \ldots, w_l\}$, where $l$ is the dimensionality of a common subspace shared by all the domains.

The goal of our method is to obtain a good clustering performance for the target domain in the common subspace. To achieve it, we are going to appropriately transfer the constraint knowledge from multiple source domains $D_s$, $s = 1, 2, \ldots, P$, to the target domain $D_T$.

### 2.1 Motivation for Transferred Centroid Regularization

In this subsection, we show a motivating example that empirically explains why we build a transferred centroid regularization in a common subspace where the difference in distributions among domains is reduced. We follow the way of [14] to generate five domains, where the first four domains are regarded as the source domains while the last domain is the target domain, as shown from Fig. 1a to Fig. 1e. Each domain consists of examples belonging to one of two classes. The examples of class one (denoted with stars) are drawn from a Gaussian Mixture Model (GMM) and the examples of class two (denoted with circles) are drawn from a single Gaussian. For the domains 1 and 2, the GMM parameters of class one are prescribed by a three-component mixture defined as follows. The mixture weights are $(0.3, 0.3, 0.4)$; their respective means are $(1, 1), (3, 3)$ and $(5, 5)$; their respective covariance matrices are $\Sigma_1 = \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 3.0 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 3.0 & 0.0 \\ 0.0 & 0.3 \end{pmatrix}$ and $\Sigma_3 = \begin{pmatrix} 3.0 & -0.5 \\ -0.5 & 0.3 \end{pmatrix}$. The examples of class two are drawn from a single Gaussian with mean $(3.5, 0.5)$ and diagonal covariance with symmetric variance 0.5. For the domains 3, 4 and the target domain, the three component GMM parameters of class one are prescribed as follows. The mixture weights are $(0.3, 0.3, 0.4)$; their respective means are $(1, 1), (3, 3)$ and $(5, 5)$; their respective covariance matrices are $\Sigma_2, \Sigma_3$ and $\Sigma_1$. The examples of class two are drawn from a single Gaussian with mean $(0.5, 3.5)$ and diagonal covariance with symmetric variance 0.5. Note that each source domain has 50 must-link constraints and 50 cannot-link constraints, which are generated randomly. To keep the
Fig. 1. Motivating example

Figures readable, the pairwise constraints are not shown in the Figures. This motivating example is designed to obtain a one-dimensional common subspace where the clustering for the target domain is performed.

From Fig. 1a to Fig. 1e, we can see that the data between domains 1 and 2 are similar and the data among domains 3, 4 and 5 are similar. However, the distributions between two groups of domains are obviously different. In Fig. 1f, we show that the similarities among ten centroids in the common subspace from all the source domains and the target domain in Hinton diagram [11], in which the square size of the \((i, j)\)-th element denotes the degree of similarity between the \(i\)-th and \(j\)-th centroids. The similarity is computed by \(\exp(-\|f_i - f_j\|^2/2)\), where \(f_i\) represents the \(i\)-th centroid. Note that a larger square in the Hinton diagram indicates a larger value of the similarity between two centroids. It is obvious from