Algorithm 1 Expectation-Maximization Data-centric (EM-DC) over $\mathcal{D}$

INPUT $\mathcal{D}$: data, k: cluster number.
OUTPUT $\mathcal{C}_1, \ldots, \mathcal{C}_k$: k-normal distributions/clusters.
// Each $\mathcal{C}_j^i \sim \mathcal{N}(\mu_j^i, \Sigma_j^i)$, $P(\mathcal{C}_j^i), w_{\mathbf{x}_j}^i, X_j^i \in \mathcal{C}_j^i$.
// i: iteration number, j:cluster number.
// μ_j^i: mean, Σ_j^i: covariance, $P(\mathcal{C}_j^i)$: prior.
// $w_{\mathbf{x}_j}^i \in \mathbb{R}$: likelihood and $X_j^i \subseteq \mathcal{D}$.
// Binary Search Trees (BSTs): $\mathbf{T}^i = \{\mathcal{T}_1^i, \mathcal{T}_2^i, \ldots, \mathcal{T}_k^i\}, \cup_{t=1}^k \mathcal{T}_t = \mathcal{D}$
randomly construct $\mathbf{C}^0 = \{\mathcal{C}_1^0, \mathcal{C}_2^0, \ldots, \mathcal{C}_k^0\}$
$i \leftarrow 0$
// $\mathcal{D}_{HE}$: high expressive data. All data points are high expressive at the first iteration.
$\mathcal{D}_{HE} \leftarrow \mathcal{D}$
repeat
    for $\mathbf{x} \in \mathcal{D}$ **do**
        // E-step:
        for $\mathcal{C}_j^i \in \mathbf{C}^i$ **do**
            $\mathcal{C}_j^i.w_{\mathbf{x}_j}^i \leftarrow P(\mathcal{C}_j^i \mid \mathbf{x})$
        end for
        // If $i = 0$, build BSTs, otherwise, update them–depends on $w_{\mathbf{x}_j}^i$.
        $\mathcal{C}_j^i.\mathcal{T}_j^i.insert(\mathbf{x}, w_{\mathbf{x}_j}^i) \leftarrow (\mathbf{x}, w_{\mathbf{x}_j}^i)$
    end for
    // $\mathcal{D}_{HE}$: A temporary variable used to store the HE data at each iteration.
    $\mathcal{D}'_{HE} \leftarrow \varnothing$
    // M-step:
    for $\mathcal{C}_j^i \in \mathbf{C}^i$ **do**
        $\mathcal{C}_j^{i+1}.\mu_j^i \leftarrow \Sigma_{\mathbf{x} \in \mathcal{C}_j^i.X_j^i}(\mathbf{x} \cdot \mathcal{C}_j^i.w_{\mathbf{x}_j}^i / (\Sigma \mathcal{C}_j^i.w_{\mathbf{x}_j}^i))$
        $\mathcal{C}_j^{i+1}.\Sigma_j^i \leftarrow \Sigma_{\mathbf{x} \in \mathcal{C}_j^i.X_j^i}(\mathcal{C}_j^i.w_{\mathbf{x}_j}^i(\mathbf{x}\text{-}\mathcal{C}_j^i.\mu_j^i)(\mathbf{x}\text{-}\mathcal{C}_j^i.\mu_j^i)^T \;/(\Sigma_j^i \mathcal{C}_j^i.w_{\mathbf{x}_j}^i))$
        $\mathcal{C}_j^{i+1}.P(\mathcal{C}_j^i) \leftarrow \Sigma(\mathcal{C}_j^i.w_{\mathbf{x}_j}^i)/|\mathcal{C}_j^i.X_j^i|$
        // Using BSTs to determine new HE data and storing them in $\mathcal{D}'_{HE}$
        // $\mathcal{C}_j^i.\mathcal{T}_j^i.flush(\Sigma)$ represents separation of HE data from BST
        $\mathcal{D}'_{HE} \leftarrow \mathcal{C}_j^i.\mathcal{T}_j^i.flush(\Sigma)$
        $\mathbf{C}^{i+1} \leftarrow \cup\{\mathcal{C}_j^{i+1}\}$
    end for
    $i \leftarrow i + 1$
    // Updating final HE data
    $\mathcal{D}_{HE} \leftarrow \mathcal{D}'_{HE}$
    // $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, Stopping criterion (Convergence over structure, BSTs):
    // Node-wise hamming distance among BSTs between two consecutive iterations.
until threshold on $d(\mathbf{C}^{i-1}, \mathbf{C}^i) \leqslant \epsilon$