Algorithm 1 Fit a Kernel Density Network to the Data

Input: (1) $(\mathbf{X}, \mathbf{y}) \in \mathbb{R}^{n \times d} \times \mathcal{Y}^n$ where $\mathcal{Y} = \{1, \ldots, K\}$

Output: (1) Gaussian mixture means μ of each class, Gaussian mixture covariances, (2) Σ of each class

function FIT($\mathbf{X}, \mathbf{y}$)
  $\theta \leftarrow$ NN.FIT($\mathbf{X}, \mathbf{y}$) ▷ $\theta = \{\mathbf{W}_i, \mathbf{b}_i\}_{i=1}^{L}$ is the set of parameters of the NN with L layers
  Let μ, Σ be k-length arrays
  for $k = 1, \ldots, K$ **do**
    Let μ_k, Σ_k be empty arrays
    Let $\mathbf{X}_k \in \mathbb{R}^{n_k \times d}, (n_k < n)$ be the matrix of data samples with class label k
    $\mathcal{P}_k \leftarrow$ GETPOLYTOPES($\mathbf{X}_k$) ▷ $\mathcal{P}_k = (\mathcal{P}_k^{(1)}, \mathcal{P}_k^{(2)}, \ldots, \mathcal{P}_k^{(|\mathcal{P}_k|)})$ contains the polytope IDs of class label k
    for $l = 1, \ldots, |\mathcal{P}_k|$ **do**
      Let $\mathbf{X}_k^{(l)} \in \mathbb{R}^{n_l \times d}, (n_l < n_k)$ be the matrix of data samples with class label k that belong to $\mathcal{P}_k^{(l)}$
      if $n_l = 0$ **then**
        continue
      end if
      $\mu_k^{(l)}, \Sigma_k^{(l)} \leftarrow$ GM.FIT($\mathbf{X}_k^{(l)}$) ▷ fit a Gaussian Mixture over $\mathbf{X}_k^{(l)}$
      μ_k.INSERT($\mu_k^{(l)}$)
      Σ_k.INSERT($\Sigma_k^{(l)}$)
    end for
    $\mu[k] \leftarrow \mu_k$
    $\Sigma[k] \leftarrow \Sigma_k$
  end for
  return (μ, Σ)
end function

Algorithm 2 Inference using the Kernel Density Network

Input: (1) $(\mathbf{X}, \mathbf{y}) \in \mathbb{R}^{n \times d} \times \mathcal{Y}^n$ where $\mathcal{Y} = \{1, \ldots, K\}$
Output: (1) Gaussian mixture means μ of each class, Gaussian mixture covariances, (2) Σ of each class

function $\textsc{fit}(\mathbf{X}, \mathbf{y})$
  $\theta \leftarrow \text{NN}.\textsc{fit}(\mathbf{X}, \mathbf{y})$ ▷ $\theta = \{\mathbf{W}_i, \mathbf{b}_i\}_{i=1}^{L}$ is the set of parameters of the NN with L layers
  Let μ, Σ be k-length arrays
  for $k = 1, \ldots, K$ **do**
    Let μ_k, Σ_k be empty arrays
    Let $\mathbf{X}_k \in \mathbb{R}^{n_k \times d}, (n_k < n)$ be the matrix of data samples with class label k
    $\mathcal{P}_k \leftarrow \textsc{getPolytopes}(\mathbf{X}_k)$ ▷ $\mathcal{P}_k = (\mathcal{P}_k^{(1)}, \mathcal{P}_k^{(2)}, \ldots, \mathcal{P}_k^{(|\mathcal{P}_k|)})$ contains the polytope IDs of class label k
    for $l = 1, \ldots, |\mathcal{P}_k|$ **do**
      Let $\mathbf{X}_k^{(l)} \in \mathbb{R}^{n_l \times d}, (n_l < n_k)$ be the matrix of data samples with class label k that belong to $\mathcal{P}_k^{(l)}$
      if $n_l = 0$ **then**
        continue
      end if
      $\mu_k^{(l)}, \Sigma_k^{(l)} \leftarrow \text{GM}.\textsc{fit}(\mathbf{X}_k^{(l)})$ ▷ fit a Gaussian Mixture over $\mathbf{X}_k^{(l)}$
      $\mu_k.\textsc{insert}(\mu_k^{(l)})$
      $\Sigma_k.\textsc{insert}(\Sigma_k^{(l)})$
    end for
    $\mu[k] \leftarrow \mu_k$
    $\Sigma[k] \leftarrow \Sigma_k$
  end for
  return (μ, Σ)
end function