DYSON: Dynamic Feature Space Self-Organization for Online Task-Free Class Incremental Learning

Anonymous CVPR submission

Paper ID 11291

Appendices

A. The proof of Theorem 1

Theorem 1. *During incremental learning, the DNC algorithm extends a simplex ETF with K prototypes $\mathbf{Z}$ to one with $K+C$ prototypes $\mathbf{Z}'$ without loss of the geometry optimality,* i.e.*, the prototypes in $\mathbf{Z}'$ satisfy:*

$$\forall i,j,\ \mathbf{z}^{i'T}\mathbf{z}^{j'} = \frac{K+C}{K+C-1}\delta_{i,j} - \frac{1}{K+C-1}, \tag{1}$$

where $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ for the opposite.

Proof. Taking the $K+C$ class centers collected in $\mathcal{P}=\{\mathbf{p}^1,...,\mathbf{p}^{K+C}\}$, where each element $\mathbf{p}^i\in\mathbb{R}^{d\times 1}$ denote the within-class mean of the features of class i. Following the proposed DNC algorithm, we first concatenate the class centers into a matrix $\mathbf{P}'=[\mathbf{p}^1,...,\mathbf{p}^{K+C}]\in\mathbb{R}^{d\times(K+C)}$, and compute the QR-decomposition of the matrix $\mathbf{P}'$:

$$\mathbf{P}' = \mathbf{Q}'\mathbf{R}', \tag{2}$$

where $\mathbf{Q}'=[\mathbf{q}^1,...,\mathbf{q}^{K+C}]\in\mathbb{R}^{d\times(K+C)}$ is an column-wise orthogonal matrix with $\mathbf{Q}^T\mathbf{Q}=\mathbf{I}_{K+C}$, *i.e.*, $\forall i\neq j, \mathbf{q}^{i^T}\mathbf{q}^j=0$ and $\forall i, \mathbf{q}^{i^T}\mathbf{q}^i=1$, and $\mathbf{R}\in\mathbb{R}^{(K+C)\times(K+C)}$ is an upper triangular matrix. We then compute the class prototypes of the $(K+C)$ classes by:

$$\mathbf{Z}' = \sqrt{\frac{K+C}{K+C-1}}\mathbf{Q}'(\mathbf{I}_{K+C} - \frac{1}{K+C}\mathbf{1}_{K+C}\mathbf{1}_{K+C}^T), \tag{3}$$

where $\mathbf{Z}=[\mathbf{z}^1,...,\mathbf{z}^{K+C}]\in\mathbb{R}^{d\times(K+C)}$ is the class prototype matrix and each column $\mathbf{z}^i\in\mathbb{R}^{d\times 1}$ denotes the class prototype of class i. For $\mathbf{z}^i$, we have:

$$\mathbf{z}^{i'} = \sqrt{\frac{K+C}{K+C-1}}\cdot(\mathbf{q}^i - \frac{\sum_{n=1}^{K+C}\mathbf{q}^n}{K+C}). \tag{4}$$

Then, for any class prototypes $\mathbf{z}^{i'}$ and $\mathbf{z}^{j'}$ in $\mathbf{Z}'$,

$$\begin{aligned}\forall i,j,\ \ \mathbf{z}^{i'T}\mathbf{z}^{j'} &= \frac{K+C}{K+C-1}\cdot(\mathbf{q}^i - \frac{\sum_{n=1}^{K+C}\mathbf{q}^n}{K+C})^T \\ &\qquad \cdot(\mathbf{q}^j - \frac{\sum_{n=1}^{K+C}\mathbf{q}^n}{K+C}), \\ &= \frac{K+C}{K+C-1}\cdot(\mathbf{q}^{i^T}\mathbf{q}^j - \frac{1}{K+C}), \\ &= \frac{K+C}{K+C-1}\cdot\delta_{i,j} - \frac{1}{K+C-1}.\end{aligned} \tag{5}$$

where $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ if $i\neq j$. On this basis, the $K+C$ prototypes construct an optimal geometry that all the prototypes are unit vectors, *i.e.*, $\forall i, ||\mathbf{z}^{i'}||_2=\mathbf{z}^{i'T}\mathbf{z}^{i'}=1$, and maximally separates the feature space using the $K+C$ vectors with the same pair-wise angles, *i.e.*, $\forall i\neq j, ||\mathbf{z}^{i'}||_2=\mathbf{z}^{i'T}\mathbf{z}^{ij'}=-\frac{1}{K+C-1}$. Thus, the $K+C$ prototypes in $\mathbf{Z}'$ computed by the DNC algorithm construct an optimal geometry structure classifying $K+C$ classes.

B. The proof of Theorem 2

Theorem 2. *With M prototype placeholders, extending existing K-class prototypes to $K+C$ ones using the DNC algorithm when C new class emerged, the upper bound of the old class prototypes is guaranteed, i.e., $\forall i=1,...,K, ||\mathbf{z}^i-\mathbf{z}^{i'}||^2\leq 2-2\sqrt{\frac{(M-1)(M+C)}{M(M+C-1)}}$, where $\mathbf{z}^i$ and $\mathbf{z}^{i'}$ are the old and new prototypes of class i, respectively.*

Proof. For any old class i, *i.e.*, $\forall i=1,...,K$, its shift distance can be computed by:

$$\begin{aligned}||\mathbf{z}^i-\mathbf{z}^{i'}||^2 &= ||\mathbf{z}^i||^2+||\mathbf{z}^{i'}||^2-2\,||\mathbf{z}^i||\cdot||\mathbf{z}^{i'}||\cdot cos\theta, \\ &= 2-2cos\theta,\end{aligned} \tag{6}$$

where

$$
\begin{aligned}
cos\theta &= \frac{\mathbf{z}^{i^T}\mathbf{z}^{i'}}{||\mathbf{z}^i|| \cdot ||\mathbf{z}^{i'}||} = \mathbf{z}^{i^T}\mathbf{z}^{i'}, \\
&= \sqrt{\frac{K}{K-1}}(\mathbf{q}^i - \frac{\sum \mathbf{q}^n}{K})^T \cdot \sqrt{\frac{K+C}{K+C-1}}(\mathbf{q}^i - \frac{\sum \mathbf{q}^m}{K+C}), \\
&= \sqrt{\frac{K(K+C)}{(K-1)(K+C-1)}}(\mathbf{q}^{i^T}\mathbf{q}^i - \frac{\sum \mathbf{q}^{n^T}\mathbf{q}^i}{K+C} - \\
&\qquad \frac{\mathbf{q}^{i^T}\sum \mathbf{q}^m}{K} + \frac{\sum \mathbf{q}^n \sum \mathbf{q}^m}{K(K+C)}), \\
&= \sqrt{\frac{K(K+C)}{(K-1)(K+C-1)}}(1 - \frac{1}{K+C} - \\
&\qquad \frac{1}{K} + \frac{K}{K(K+C)}), \\
&= \sqrt{\frac{K(K+C)}{(K-1)(K+C-1)}} \cdot \frac{K-1}{K}, \\
&= \sqrt{\frac{(K-1)(K+C)}{K(K+C-1)}}.
\end{aligned}
\tag{7}
$$

Then, we have

$$
||\mathbf{z}^i - \mathbf{z}^{i'}||^2 = 2 - 2\sqrt{\frac{(K-1)(K+C)}{K(K+C-1)}}. \tag{8}
$$

When $C \geq 1$, Eq. (8) is a monotonical decreasing function *w.r.t* the increase of K. By adding M placehoders in our DNC algorithm, we have $K \geq M$, the upper bound of the drift distance can be guaranteed, *i.e.*, $||\mathbf{z}^i - \mathbf{z}^{i'}||^2 \leq 2 - 2\sqrt{\frac{(M-1)(M+C)}{M(M+C-1)}}$.