Assignment 6: Extension Proof

Advanced Topics In Machine Learning (CS6360)

Rahul Vigneswaran

CS23MTECH02002

Paper presented as part of assignment 1

Neural Collapse Terminus: A Unified Solution for Class Incremental Learning and Its Variants

We consider the following problem,

$$\min_{\mathbf{M}^{(t)}} \quad \frac{1}{N^{(t)}} \sum_{k=1}^{K^{(t)}} \sum_{i=1}^{n_k} \mathcal{L}\left(\mathbf{m}_{k,i}^{(t)}, \hat{\mathbf{W}}_{\mathrm{ETF}}\right),\ 0 \le t \le T,$$

$$s.t. \quad \|\mathbf{m}_{k,i}^{(t)}\|^2 \le 1, \quad \forall 1 \le k \le K^{(t)},\ 1 \le i \le n_k,$$

where,

$\mathbf{m}_{k,i}^{(t)} \in \mathbb{R}^d$: i-th sample of class k in session t feature
n_k : no. of samples in class k
$K^{(t)}$: no. of classes in session t
$N^{(t)} = \sum_{k=1}^{K^{(t)}} n_k$
$\mathbf{M}^{(t)} \in \mathbb{R}^{d \times N^{(t)}}$: collection of $\mathbf{m}_{k,i}^{(t)}$
$K = \sum_{t=0}^{T} K^{(t)}$
$\hat{\mathbf{W}}_{\mathrm{ETF}} \in \mathbb{R}^{d \times K}$: neural collapse terminus all K

A simplex equiangular tight frame (ETF) refers to a collection of vectors $\{\mathbf{e}_i\}_{i=1}^{K}$ in $\mathbb{R}^d$, $d \geq K - 1$, that satisfies:

$$\mathbf{e}_{k_1}^T \mathbf{e}_{k_2} = \frac{K}{K-1}\delta_{k_1,k_2} - \frac{1}{K-1}, \qquad \text{DEF1}$$

$$\forall k_1, k_2 \in [1, K],$$

where $\delta_{k_1,k_2} = 1$ when $k_1 = k_2$, and 0 otherwise. All vectors have the same ℓ_2 norm and any pair of two different vectors has the same inner product of $-\frac{1}{K-1}$, which is the minimum possible cosine similarity for K equiangular vectors in $\mathbb{R}^d$.

Theorem

Let $\hat{\mathbf{M}}^{(t)}$ denotes the global minimizer by optimizing the model incrementally from $t = 0$, and we have $\hat{\mathbf{M}} = [\hat{\mathbf{M}}^{(0)}, \cdots, \hat{\mathbf{M}}^{(T)}] \in \mathbb{R}^{d \times \sum_{t=0}^{T} N^{(t)}}$. No matter if $\mathcal{L}$ is CE or misalignment loss, for any column vector $\hat{\mathbf{m}}_{k,i}$ in $\hat{\mathbf{M}}$ whose class label is k, we have:

$$\|\hat{\mathbf{m}}_{k,i}\| = 1, \quad \hat{\mathbf{m}}_{k,i}^T \hat{\mathbf{w}}_{k'} = \frac{K}{K-1}\delta_{k,k'} - \frac{1}{K-1},$$

for all $k, k' \in [1, K]$, $1 \leq i \leq n_k$, where $K = \sum_{t=0}^{T} K^{(t)}$ denotes the total number of classes of the whole label space, $\delta_{k,k'} = 1$ when $k = k'$ and 0 otherwise, and $\hat{\mathbf{w}}_{k'}$ is the class prototype in $\hat{\mathbf{W}}_{\mathrm{ETF}}$ for class k'.

Extension: Relaxing implicit weight assumption

$$\min_{\mathbf{M}^{(t)}} \quad \frac{1}{N^{(t)}} \sum_{k=1}^{K^{(t)}} \sum_{i=1}^{n_k} \mathcal{L}\left(\mathbf{m}_{k,i}^{(t)}, \hat{\mathbf{W}}_{\mathrm{ETF}}\right), \ 0 \leq t \leq T,$$
$$s.t. \quad \|\mathbf{m}_{k,i}^{(t)}\|^2 \leq 1, \quad \forall 1 \leq k \leq K^{(t)}, \ 1 \leq i \leq n_k,$$

$$\min_{\mathbf{M}^{(t)}} \quad \sum_{k=1}^{K^{(t)}} \frac{1}{N_k^{(t)}} \sum_{i=1}^{n_k} \mathcal{L}\left(\mathbf{m}_{k,i}^{(t)}, \hat{\mathbf{W}}_{\mathrm{ETF}}\right), \ 0 \leq t \leq T,$$
$$s.t. \quad \|\mathbf{m}_{k,i}^{(t)}\|^2 \leq 1, \quad \forall 1 \leq k \leq K^{(t)}, \ 1 \leq i \leq n_k,$$

Lagrangian Function

$$\tilde{L} = \sum_{k=1}^{K} \alpha_k \sum_{i=1}^{n_k} -\log\left(\frac{\exp\left(\hat{W}_k^T m_{k,i}\right)}{\sum_{j=1}^{K} \exp\left(\hat{W}_j^T m_{k,i}\right)}\right) + \sum_{k=1}^{K} \sum_{i=1}^{n_k} \lambda_{k,i}\left(||m_{k,i}||^2 - 1\right)$$

Derivative

$$\frac{\partial \tilde{L}}{\partial m_{k,i}} = -\alpha_k(1 - p_k)w_k + \alpha_k \sum_{j \neq k}^{K} p_j w_j + 2\lambda_{k,i} m_{k,i}$$

$\lambda = 0$

$$-\alpha_k(1-p_j)w_k + \alpha_k \sum_{j\neq k}^{K} p_j w_j = 0$$

$$(1-p_j)w_k = \sum_{j\neq k}^{K} p_j w_j$$

Multiply by w_k,

$$(1-p_j)w_k^T w_k = \sum_{j\neq k}^{K} p_j w_j^T w_k$$

$$(1-p_k)\left(\frac{K}{K-1}\right) = 0$$

$$\implies p_k = 1 \qquad \text{Contradiction.}$$

$\lambda = 0$

Based on KKT's complimentary slackness condition,

$$\lambda_{k,i}\left(||m_{k,i}||^2 - 1\right) = 0$$

$$\implies ||m_{k,i}|| = 1 \qquad \text{EQN0.1}$$

$$\mathbf{W}_{ETF} \cdot \mathbf{1}_k = \mathbf{0}_d$$

$$\sum_{k=1}^{K} w_k = 0_d \qquad \text{EQN0.2}$$

$\lambda > 0$

$$-\alpha_k(1 - p_k)w_k + \alpha_k \sum_{j \neq k}^{K} p_j w_j + 2\lambda_{k,i} m_{k,i} = 0$$

$$-\alpha_k(\sum_{j \neq k}^{K} p_j)w_k + \alpha_k \sum_{j \neq k}^{K} p_j w_j + 2\lambda_{k,i} m_{k,i} = 0$$

$$\alpha_k \sum_{j \neq k}^{K} p_j(w_j - w_k) + 2\lambda_{k,i} m_{k,i} = 0 \qquad \text{EQN1}$$

$\lambda > 0$

EQN1 $\times\, w_{j'}, \forall j' \neq k$,

$$\alpha_k \sum_{j \neq k}^{K} p_j w_j^T w_{j'} - \alpha_k \sum_{j \neq k}^{K} p_j w_k^T w_{j'} + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$\alpha_k p_{j'} w_{j'}^T w_{j'} + \alpha_k \sum_{\substack{j \neq k \\ j \neq j'}}^{K} p_j w_j^T w_{j'} - \alpha_k \sum_{j \neq k}^{K} p_j w_k^T w_{j'} + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$\alpha_k p_{j'} + \alpha_k \left(\frac{-1}{K-1} \right) \sum_{\substack{j \neq k \\ j \neq j'}}^{K} p_j - \alpha_k \left(\frac{-1}{K-1} \right) \sum_{j \neq k}^{K} p_j + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$\alpha_k p_{j'} - \frac{\alpha_k}{K-1} \left(1 - p_k - p_{j'}\right) p_j + \left(\frac{\alpha_k}{K-1} \right) \left(1 - p_k\right) + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$\alpha_k p_{j'} \frac{K}{K-1} + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

Prove $p_{j_1} = p_{j_2}$

$$p_{j'} = -\frac{2\lambda_{k,i} m_{k,i}^T w_{j'} (K-1)}{\alpha_k K}$$

$$\frac{p_{j_1}}{p_{j_2}} = \frac{m_{k,i}^T w_{j_1}}{m_{k,i}^T w_{j_2}} = \frac{\exp\left(m_{k,i}^T w_{j_1}\right)}{\exp\left(m_{k,i}^T w_{j_2}\right)}$$

$$\frac{p_{j_1}}{p_{j_2}} = \frac{\exp\left(m_{k,i}^T w_{j_2}\right)}{m_{k,i}^T w_{j_1}} = \frac{\exp\left(m_{k,i}^T w_{j_1}\right)}{m_{k,i}^T w_{j_2}}$$

The function $f(x) = \frac{\exp(x)}{x}$ is monotonically decreasing when $x < 1$.
So, $p_{j_1} = p_{j_2}, m_{k,i}^T w_{j_1} = m_{k,i}^T w_{j_2}, \forall j_1, j_2 \neq k$

$p_{j_1} = p_{j_2}$

Since $p_{j_1} = p_{j_2}, \forall j_1, j_2 \neq k$,

$$\sum_{j \neq k} p_j = (1 - p_k)$$

$$(K-1)p_j = (1 - p_k)$$

$$p_j = \frac{1 - p_k}{K - 1} \qquad \text{EQN2.1}$$

$\lambda > 0$

Now lets rewrite EQN2 based on EQN2.1,

$$\alpha_k p_{j'} \frac{K}{K-1} + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$\alpha_k \left(\frac{1-p_k}{K-1}\right)\left(\frac{K}{K-1}\right) + 2\lambda_{k,i} m_{k,i}^T w_{j'} = 0$$

$$-\alpha_k \frac{K(1-p_k)}{(k-1)^2} = 2\lambda_{k,i} m_{k,i}^T w_{j'} \qquad \text{EQN3}$$

$\lambda > 0$

EQN1 $\times$ w_k,

$$\alpha_k \sum_{j\neq k} p_j {w_j}^T w_k - \alpha_k \sum_{j\neq k} p_j {w_k}^T w_k + 2\lambda_{k,i} m_{k,i}^T w_k = 0$$

$$\alpha_k \left(\frac{-1}{K-1}\right)(1-p_k) - \alpha_k(1-p_k) + 2\lambda_{k,i} m_{k,i}^T w_k = 0$$

$$\frac{\alpha_k K(1-p_k)}{K-1} = 2\lambda_{k,i} m_{k,i}^T w_k$$

Divide both sides by $-(k-1)$,

$$-\frac{\alpha_k K(1-p_k)}{(K-1)^2} = \frac{2\lambda_{k,i} m_{k,i}^T w_k}{(K-1)} \qquad \text{EQN4}$$

$\lambda > 0$

Combine EQN3 and EQN4,

$$2\lambda_{k,i} m_{k,i}^T w_{j'} = \frac{2\lambda_{k,i} m_{k,i}^T w_k}{(K-1)}$$

$$m_{k,i}^T w_{j'}(K-1) + m_{k,i}^T w_k = 0 \qquad \text{EQN5}$$

$\lambda > 0$

We already know,

$$\sum_{k}^{K} w_k = 0_d \qquad \text{EQN0.2}$$

Substituting EQN5.1 in EQN1,

$$\left[\frac{\alpha_k(1-p_k)K}{K-1}\right] w_k = \left[2\lambda_{k,i}\right] m_{k,i}$$

Hence, $m_{k,i}$ is aligned with w_k. $\cos(\angle(m_{k,i}, w_k)) = 1$. We also know $||m_{k,i}|| = 1$ and $||w_k|| = 1$ from EQN0.1.

$$m_{k,i}^T w_k = ||m_{k,i}||||w_k|| \cos(\angle(m_{k,i}, w_k))$$
$$m_{k,i}^T w_k = 1 \qquad \text{EQN6}$$

$\lambda > 0$

From EQN6 and EQN5,

$$m_{k,i}^T w_{j'} = -\frac{1}{K-1}$$
$$m_{k,i}^T w_j = -\frac{1}{K-1}$$

Therefore for any column vector $m_{k,i}$ in M, we have,

$$||m_{k,i}|| = 1, m_{k,i}^T w_{k'} = \frac{K}{K-1}\delta_{k,k'} - \frac{1}{K-1}$$

Proved.