Lagrangian Dual & KKT Conditions

Presentation 2

Rahul Vigneswaran

CS23MTECH02002

Paper presented as part of presentation 1

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

Presentation Overview

Recap

We consider the following problem,

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

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

Recap

Definition (Simplex Equiangular Tight Frame)

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:

$$\mathbf{e}_{k_1}^T \mathbf{e}_{k_2} = \frac{K}{K-1}\delta_{k_1,k_2} - \frac{1}{K-1}, \ \ \forall k_1, k_2 \in [1, K],$$

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 $\mathbb{R}^d$.

Recap

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:

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

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'.

Types of Optimization

- Unconstrained
  - Easy.
  - High school math. $f' = 0$.
- Constrained
  - Hard.
  - Convert constrained to unconstrained and solve.
  - Our interest.

Constrained Optimization

$$
\begin{aligned}
\min_{x} \quad & x_1 + x_2 \\
\text{s.t.} \quad & x_1^2 + x_2^2 - 1 \leq 0
\end{aligned}
$$

Constrained to Unconstrained

$$\min_{x} \quad x_1 + x_2 + P(x_1^2 + x_2^2 - 1)$$

$$P(y) = \begin{cases} \infty & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Constrained to Unconstrained

$$\min_{x} \quad x_1 + x_2 + P(x_1^2 + x_2^2 - 1)$$

$$P(y) = u \times y$$

Constrained to Unconstrained

$$\min_{x} \quad x_1 + x_2 + P(x_1^2 + x_2^2 - 1)$$

$$P(y) = \max_{u \geq 0} \quad u \times y$$

Constrained to Unconstrained: Primal

$$
\begin{aligned}
\min_{x} \quad & x_1 + x_2 + \max_{u \geq 0} \quad u(x_1^2 + x_2^2 - 1) \\
\min_{x} \max_{u \geq 0} \quad & x_1 + x_2 + u(x_1^2 + x_2^2 - 1)
\end{aligned}
$$

Constrained to Unconstrained: Lagrangian Dual

$$\max_{u \geq 0} \min_{x} \quad \underbrace{x_1 + x_2 + u(x_1^2 + x_2^2 - 1)}_{\text{Lagrangian Dual}}$$

Lagrangian Dual: Lets formalize

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
s.t \quad & g_i(x) \leq 0 \quad \forall i \in \{1, 2, 3, 4, \dots\}
\end{aligned}
$$

Lagrangian Dual: Lets formalize

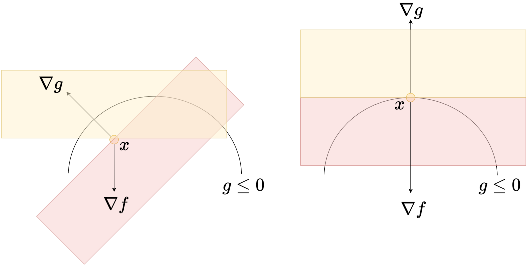
$$\min_{x} f(x) + \max_{u_i \geq 0} \sum_i u_i g_i(x)$$

$$\min_{x} \max_{u_i \geq 0} f(x) + \sum_i u_i g_i(x)$$

$$\max_{u_i \geq 0} \min_{x} f(x) + \sum_i u_i g_i(x)$$

Gradient of Lagrangian

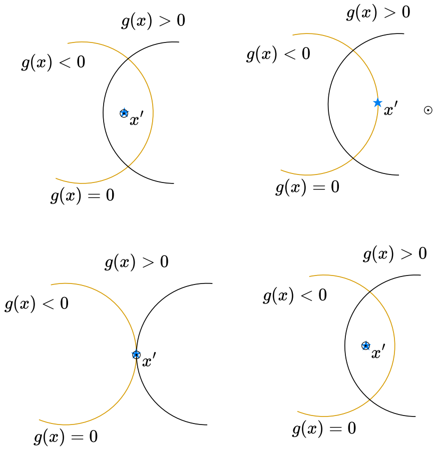
$$\nabla\left(f(x) + \sum_i u_i g_i(x)\right) = 0$$
$$\nabla f(x) + \nabla\left(\sum_i u_i g_i(x)\right) = 0$$
$$\nabla f(x) + u\nabla g(x) = 0$$
$$\nabla f(x) = -u\nabla g(x)$$



KKT Conditions

x^* and u^* are values are optimal.

1. $\nabla f(x^*) = -u^* \nabla g(x^*)$
2. $g(x^*) \leq 0$
3. $u^* \geq 0$
4. $u^* \; g(x) = 0$



Resources

- https://www.youtube.com/watch?v=uh1Dk68cfWs
- https://www.youtube.com/watch?v=s8j-pI_tPlM
- https://www-cs.stanford.edu/people/davidknowles/lagrangian_duality.pdf
- https://www.youtube.com/watch?v=3ywFQOkzTQE
- Is there a connection between lagrangian multiplier and eigenvalues? : https://journeyinmath.wordpress.com/2018/08/18/lagrangemultipliers/