lpNTK: Better Generalisation with Less Data via Sample Interaction During Learning

Shangmin Guo, Yi Ren, Stefano V. Albrecht, Kenny Smith

April 22, 2024

UKRI CENTRE FOR DOCTORAL TRAINING

lpNTK: Sample Relationship through Learning Dynamics

Samples having Similar Learning Effects

Intuition

Samples having Similar Learning Effects

Intuition




Question: how to find out the samples having similar learning effects in DL?
Answer: a novel lpNTK derived via first-Order Taylor approximation to sample interaction, i.e. how learning $\boldsymbol{x}_u$ changes the prediction on $\boldsymbol{x}_o$

Formal Definition of lpNTK

labelled pseudo neural tangent kernel (lpNTK)

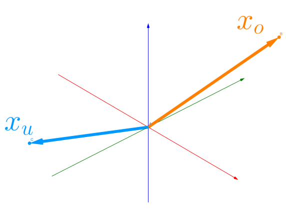
$$\begin{aligned}\kappa((\boldsymbol{x}_o, y_o), (\boldsymbol{x}_u, y_u)) &\triangleq \frac{1}{K}\sum\left[\boldsymbol{s}(y_u)\cdot\boldsymbol{s}(y_o)^{\mathsf{T}}\right]\odot\boldsymbol{K}(\boldsymbol{x}_o, \boldsymbol{x}_u)\\ &= \underbrace{\left[\frac{1}{\sqrt{K}}\boldsymbol{s}(y_o)^{\mathsf{T}}\nabla_{\boldsymbol{w}}\boldsymbol{z}(\boldsymbol{x}_o)\right]}_{1\times d}\cdot\underbrace{\left[\nabla_{\boldsymbol{w}}\boldsymbol{z}(\boldsymbol{x}_u)^{\mathsf{T}}\boldsymbol{s}(y_u)\frac{1}{\sqrt{K}}\right]}_{d\times 1}\end{aligned} \tag{1}$$

Feature representation of $(\boldsymbol{x}, y)$ under lpNTK:

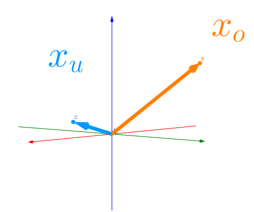
$$\frac{1}{\sqrt{K}}\boldsymbol{s}(y)^{\mathsf{T}}\nabla_{\boldsymbol{w}}\boldsymbol{z}(\boldsymbol{x}) \rightarrow \text{ a } 1\times d \text{ vector!}$$

Sample Relationships under lpNTK Feature Representation

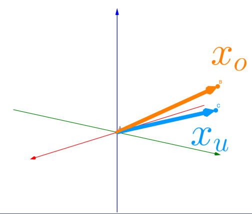
$(\boldsymbol{x}, y)$ corresponds to a vector in the gradient space



(a) Contradictory



(b) Unrelated



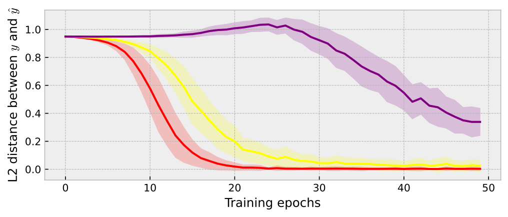
(c) Interchangeable

- Interchangeable: $\boldsymbol{x}_u \uparrow \vee \boldsymbol{x}_o \uparrow \quad \Rightarrow \quad \boldsymbol{x}_u \uparrow \wedge \boldsymbol{x}_o \uparrow$
- Unrelated: $\boldsymbol{x}_u \uparrow \Rightarrow \boldsymbol{x}_o - \quad \wedge \quad \boldsymbol{x}_o \uparrow \Rightarrow \boldsymbol{x}_u -$
- Contradictory: $\boldsymbol{x}_u \uparrow \Rightarrow \boldsymbol{x}_o \downarrow \quad \wedge \quad \boldsymbol{x}_o \uparrow \Rightarrow \boldsymbol{x}_u \downarrow$

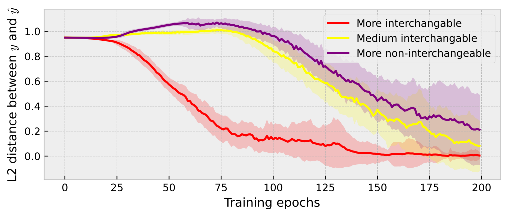
Use Case 1: Control Learning Difficulty

For a given target sample:

- More interchangeable $\rightarrow$ easier to learn
- More contradictory $\rightarrow$ harder to learn
- More unrelated $\rightarrow$ between the above two cases



(a) on MNIST



(b) on CIFAR-10

Use Case 2: Predict Forgetting Events during Learning

Predict forgetting events with lpNTK:

Benchmarks	Precision		Recall		F1-score	
	Mean	Std	Mean	Std	Mean	Std
MNIST	42.72%	$\pm 6.55\%$	59.02%	$\pm 7.49\%$	49.54%	$\pm 6.99\%$
CIFAR-10	49.47%	$\pm 7.06\%$	69.50%	$\pm 7.49\%$	57.76%	$\pm 7.36\%$

Predict forgetting events with eNTK:

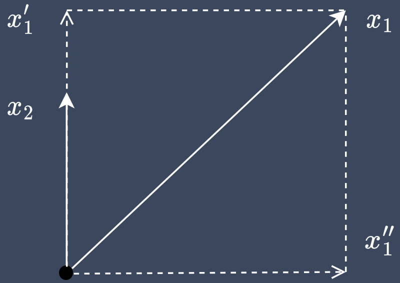
Datasets	Precision		Recall		F1-score	
	Mean	Std	Mean	Std	Mean	Std
MNIST	95.56%	$\pm 8.14\%$	86.67%	$\pm 13.67\%$	89.96%	$\pm 7.10\%$
CIFAR-10	96.99%	$\pm 3.99\%$	98.99%	$\pm 1.61\%$	97.87%	$\pm 2.08\%$

Use Case 3: Improve Generalisation Performance in Image Classification

Outline

- Do we really need all those interchangeable samples for good generalisation?
- Can we improve the generalisation performance by removing the bias in the data towards the numerous interchangeable samples?

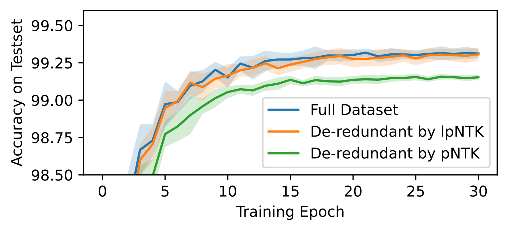
Redundant Samples



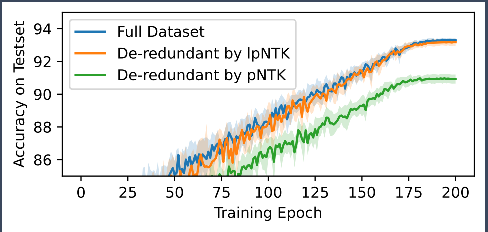
Formal Definition

For a labelled sample $(\boldsymbol{x}, y)$, if there exists another labelled sample $(\boldsymbol{x}', y')$ where $\boldsymbol{x}' \neq \boldsymbol{x}$ such that $\kappa((\boldsymbol{x}, y), (\boldsymbol{x}', y')) > \kappa((\boldsymbol{x}, y), (\boldsymbol{x}, y))$, then $(\boldsymbol{x}, y)$ is considered as a redundant sample.

Experiments on Removing Redundant Samples



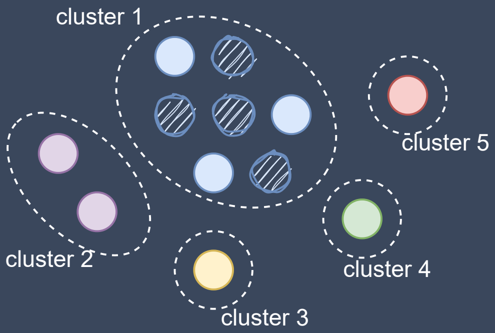
(a) MNIST



(b) CIFAR-10

Poisoning Samples

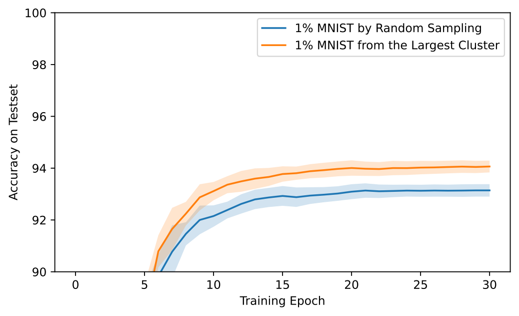
For a given set of samples $\mathbb{T}$, if performance trained on $\tilde{\mathbb{T}} > \mathbb{T}$ (where $\tilde{\mathbb{T}} \subset \mathbb{T}$), $\mathbb{T} \setminus \tilde{\mathbb{T}}$ are considered as poisoning samples.



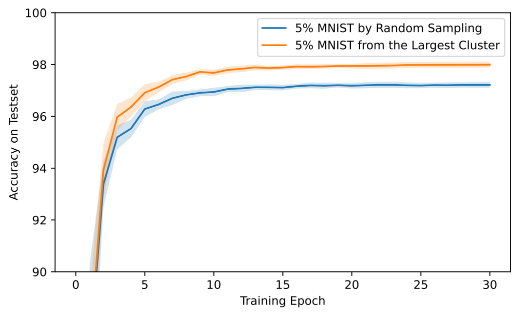
Experiment Results on Pruning Image Training Sets

Benchmarks	Full	lpNTK	EL2N	GraNd	Forgot Score
MNIST	$99.31(\pm 0.03)$%	$99.37(\pm 0.04)$%	$99.33(\pm 0.06)$%	$99.28(\pm 0.05)$%	$99.26(\pm 0.06)$%
CIFAR10	$93.28(\pm 0.06)$%	$93.55(\pm 0.12)$%	$93.32(\pm 0.07)$%	$92.87(\pm 0.13)$%	$92.64(\pm 0.22)$%

Side Point: Remove Small Clusters when #Samples is Small



(a) 1%



(b) 5%

Thank You!