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 x_u changes the prediction on x_o

Formal Definition of lpNTK

labelled pseudo neural tangent kernel (lpNTK)

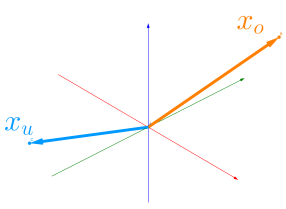
$$
\begin{aligned}
((x_o, y_o), (x_u, y_u)) &\triangleq \frac{1}{K} \sum \left[s(y_u) - s(y_o)^{\mathsf{T}} \right] K(x_o, x_u) \\
&= \underbrace{\left[\frac{1}{\sqrt{K}} s(y_o)^{\mathsf{T}} \nabla_w z(x_o) \right]}_{1 \times d} \underbrace{\left[\nabla_w z(x_u)^{\mathsf{T}} s(y_u) \frac{1}{\sqrt{K}} \right]}_{d \times 1}
\end{aligned} \tag{1}
$$

Feature representation of (x, y) under lpNTK:

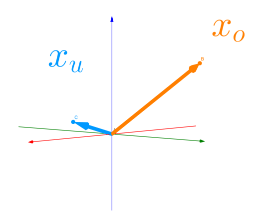
$$\frac{1}{\sqrt{K}} s(y)^{\mathsf{T}} \nabla_w z(x) \rightarrow \text{a } 1 \times d \text{ vector!}$$

Sample Relationships under lpNTK Feature Representation

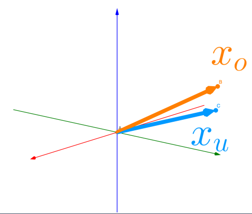
$(x; y)$ corresponds to a vector in the gradient space



(a) Contradictory



(b) Unrelated



(c) Interchangeable

Interchangeable: x_u " _x_o ") x_u " ^x_o "

Unrelated: x_u ") x_o ^ x_o ") x_u

Contradictory: x_u ") x_o # ^ x_o ") x_u #

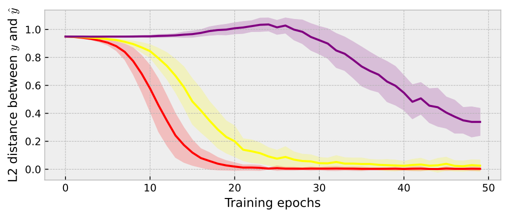
Use Case 1: Control Learning Difficulty

For a given target sample:

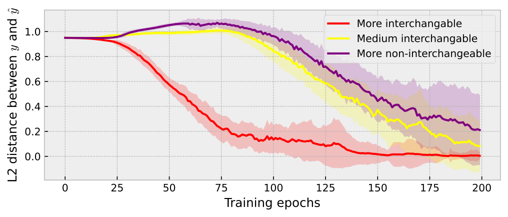
More interchangeable ! easier to learn

More contradictory ! harder to learn

More unrelated ! 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%	6:55%	59:02%	7:49%	49:54%	6:99%
CIFAR-10	49:47%	7:06%	69:50%	7:49%	57:76%	7:36%

Predict forgetting events with eNTK:

Datasets	Precision		Recall		F1-score	
	Mean	Std	Mean	Std	Mean	Std
MNIST	95:56%	8:14%	86:67%	13:67%	89:96%	7:10%
CIFAR-10	96:99%	3:99%	98:99%	1:61%	97:87%	2:08%

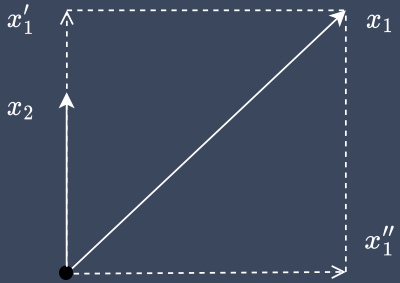
Use Case 3: Improve Generalisation Performance in Image Classification

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?

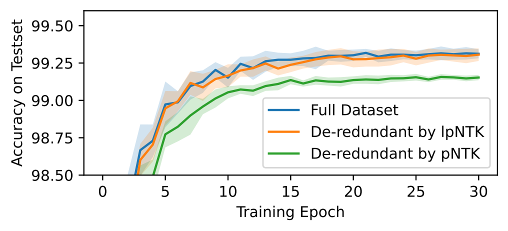
Redundant Samples



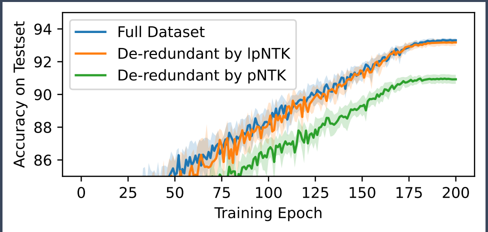
Formal Definition

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

Experiments on Removing Redundant Samples



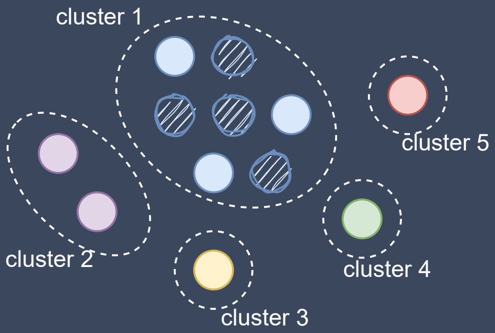
(a) MNIST



(b) CIFAR-10

Poisoning Samples

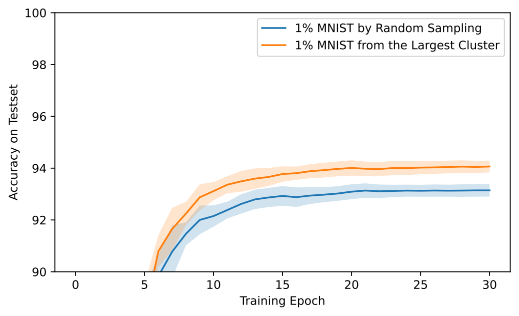
For a given set of samples $\mathbb{T}$, if performance trained on $\mathbb{T} > \mathbb{T}$ (where $\mathbb{T} \quad \mathbb{T}$), $\mathbb{T}\ n\ \mathbb{T}$ are considered as poisoning samples.



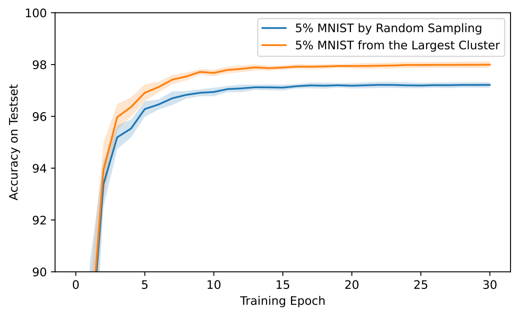
Experiment Results on Pruning Image Training Sets

Benchmarks	Full	lpNTK	EL2N	GraNd	Forgot Score
MNIST	99:31(0:03)%	99:37(0:04)%	99:33(0:06)%	99:28(0:05)%	99:26(0:06)%
CIFAR10	93:28(0:06)%	93:55(0:12)%	93:32(0:07)%	92:87(0:13)%	92:64(0:22)%

Side Point: Remove Small Clusters when #Samples is Small



(a) 1%



(b) 5%

Thank You!