Dimensionality-Driven Learning with Noisy Labels

Xingjun Ma*[1], Yisen Wang*[2], Michael E. Houle[3], Shuo Zhou[1], Sarah M. Erfani[1], Sudanthi Wijewickrema[1], James Bailey[1]

[1]The University of Melbourne, [2]Tsinghua University, [3]National Institute of Informatics, Japan

Why

- **Training deep neural networks (DNNs) robustly on data with noisy (incorrect) labels is important to deep learning.**
- **DNNs overfit to noisy labels and generalize poorly, and their learning behaviours require further understanding.**

What

- **We investigate learning behaviours of deep neural networks (DNNs) on clean labels versus noisy labels, from the view point of subspace dimensionality.**
- **We propose Dimensionality-Driven Learning (D2L) to robustly train DNNs with noisy labels.**

Measuring Subspace Dimensionality

Expansion Dimension

- Given two balls of differing radii r_1 and r_2, dimension m can be deduced from ratios of volumes:

$$\frac{V_2}{V_1} = \left(\frac{r_2}{r_1}\right)^m \Rightarrow m = \frac{\ln(V_2/V_1)}{\ln(r_2/r_1)}$$

- V_1 and V_2 are estimated by the numbers of points contained in the two balls.

Local Intrinsic Dimensionality

Given a data sample $x \in X$, let $r>0$ be a random variable denoting the distance from x to other data samples. The local intrinsic dimension of x at distance r is

$$\mathrm{LID}_F(r) \triangleq \lim_{\epsilon \to 0^+} \frac{\ln(F((1+\epsilon)\cdot r)/F(r))}{\ln(1+\epsilon)} = \frac{r \cdot F'(r)}{F(r)},$$

wherever the limit exists.

- $F(r)$: **cdf** of the distribution of distances to data from a given reference location.

Estimation of LID

- Maximum Likelihood Estimator (Hill 1975, Amsaleg et al. 2015):

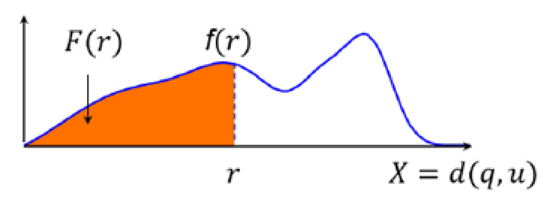
$$\widehat{\mathrm{LID}}(x) = -\left(\frac{1}{k}\sum_{i=1}^{k} \log \frac{r_i(x)}{r_k(x)}\right)^{-1}$$

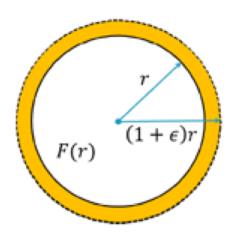
- Extreme Value Theory:
  - Nearest distances are extreme events.
  - Lower tail distribution follows Generalized Pareto Distribution.
- Efficient estimation within a random minibatch (Ma et al. 2018).

Interpretation of LID

$$\mathrm{LID}_F(r) = \frac{r \cdot F'(r)}{F(r)}$$

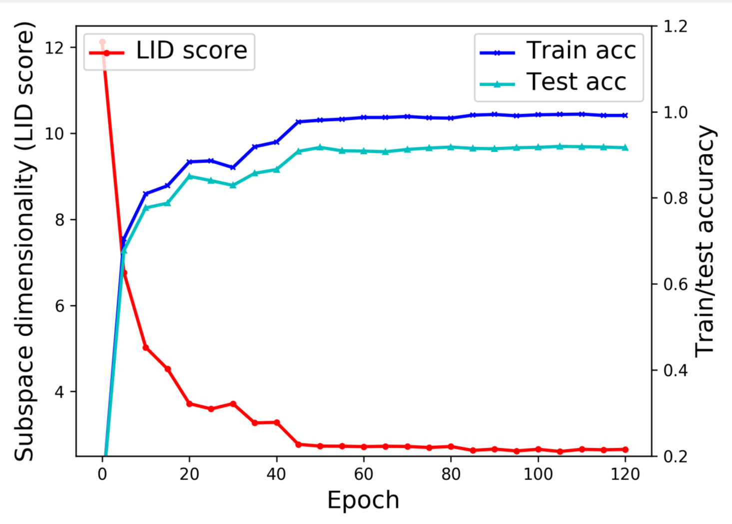
- LID_F (r): measures growth rate of $F(r)$ as the radius r expands (Houle 2017a).





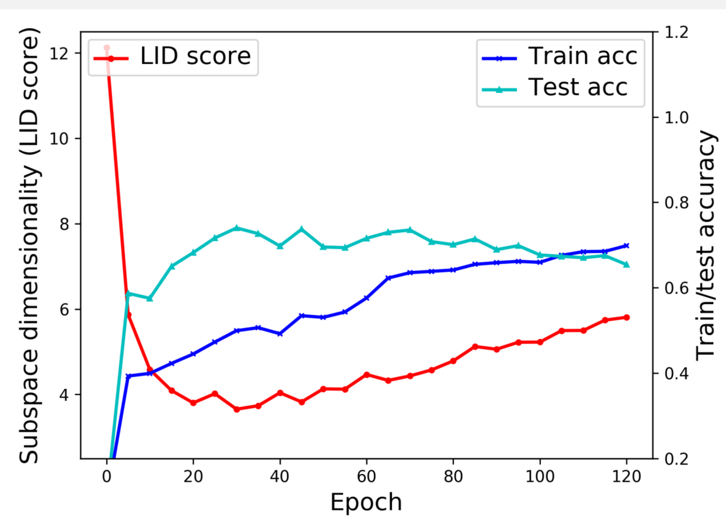
Dimensionality-Driven Learning (D2L)

Learning with Clean Labels



- **Decreasing subspace dimensionality: compression.**

Learning with Noisy Labels



- **Dimensionality shift from compression to expansion.**

Interpretation on Dimensionality Shift

- ✓ DNNs learn simple submanifolds first, then increase submanifold complexity to accommodate noisy samples.
- ✓ Noisy samples are hard to compress into low dimensional submanifolds.
- ✓ **Dimensionality shift indicates noisy learning.**

Dimensionality-Driven Learning

- Adjusted cross-entropy: questions original labels based on the degree of expansion α_i.

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{y_n^*} y_n^* \log P(y_n^*|x_n)$$

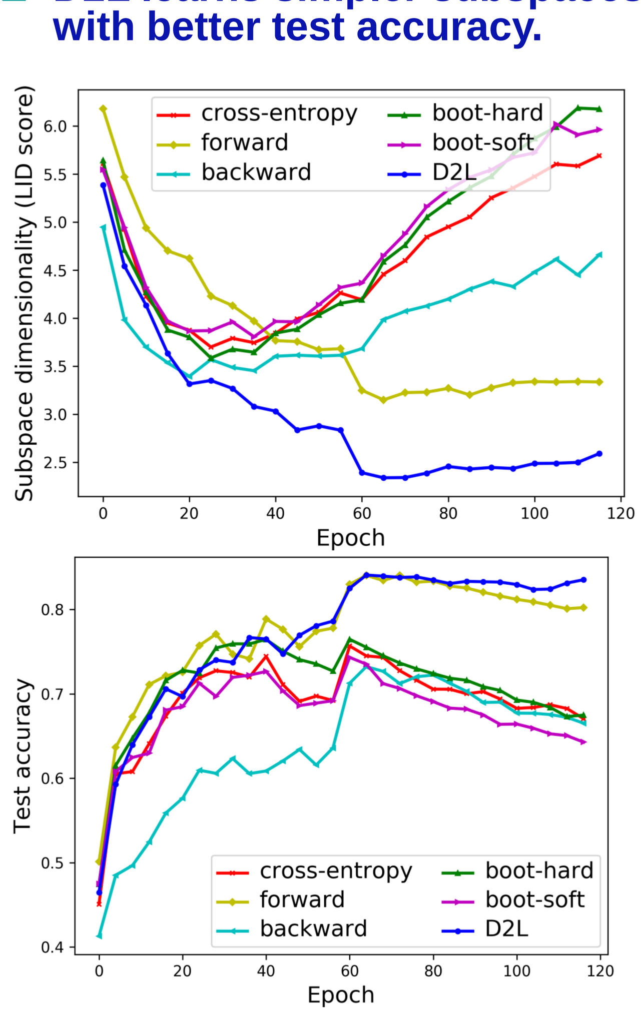
$$y^* = \alpha_i y + (1-\alpha_i)\hat{y}$$

$$\alpha_i = exp(-\lambda \frac{LID_i}{min_j^{i-1} LID_j})$$

i: current epoch, T: total number of epochs, y: original label, $\hat{y}$: predicted label. α_i: LID-based weighting for the label interpolation.

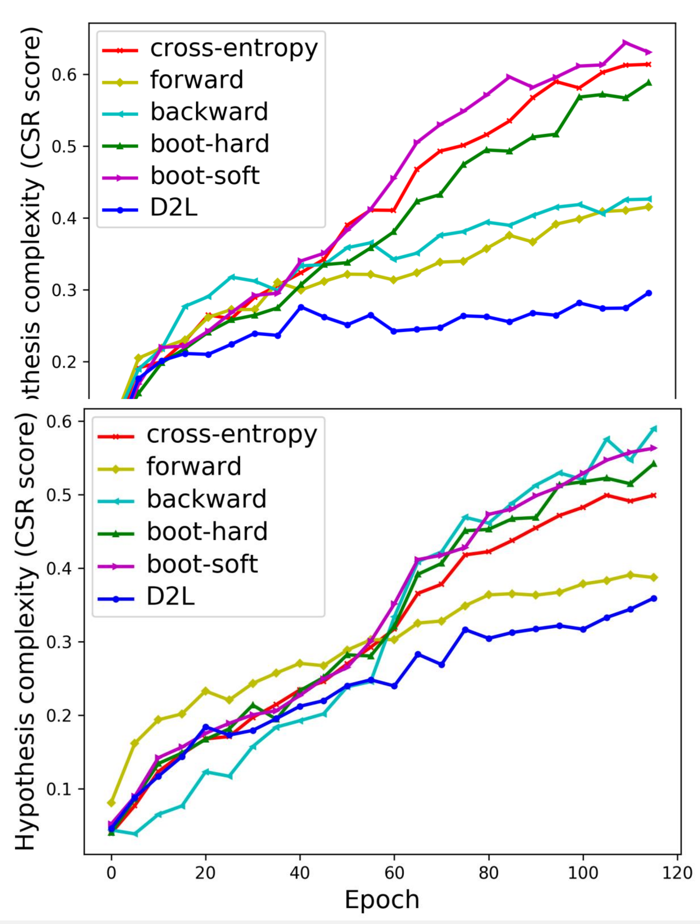
Subspace Learning

- **D2L learns simpler subspaces with better test accuracy.**



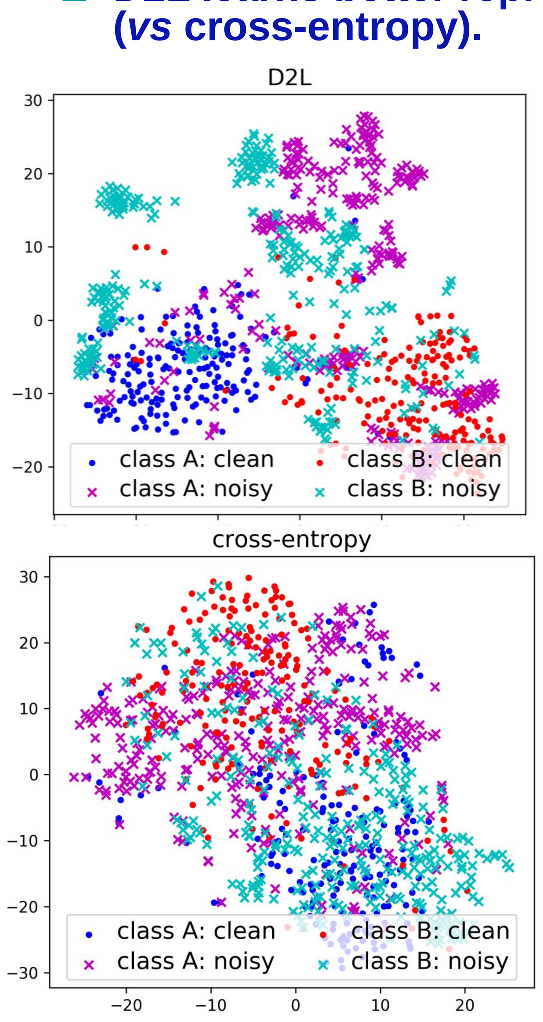
Hypothesis Learning

- **D2L learns simpler hypothesis (Critical Sample Ratio): Top: 40% Bottom:60%, random flipping.**



Representation Learning

- **D2L learns better representation (vs cross-entropy).**



CIFAR-10, 60% noise.

A: 'airplane'. Noisy A: random.

B: 'cat'. Noisy B: random.

Robustness to Noisy Labels

- MNIST: 5-layer CNN; SVHN: 6-layer CNN; CIFAR-10: 12-layer CNN; CIFAR-100: ResNet-44.
- Noise rates: 20% - 60% random flipping (symmetric noise).
- D2L delivers strong classification performance across the tested noise rates.

Table 1: Test accuracy (%) $\pm$ stdev

Dataset / Noise Rate		cross-entropy	forward	backward	boot-hard	boot-soft	D2L
MNIST	0%	99.24±0.0	**99.30±0.0**	99.23±0.1	99.13±0.2	99.20±0.0	99.28±0.0
	20%	82.66±1.8	96.45±0.4	84.69±1.2	80.69±2.2	83.50±1.2	**98.84±0.0**
	40%	60.14±3.9	88.90±0.9	64.89±1.0	60.49±1.6	59.19±1.8	**98.49±0.0**
	60%	38.51±3.7	72.88±1.6	42.83±3.3	40.45±1.6	39.04±3.0	**94.73±1.2**
SVHN	0%	90.12±0.3	90.22±0.1	90.16±0.2	89.47±0.0	89.26±0.0	**90.32±0.0**
	20%	76.10±0.9	85.51±0.7	74.61±0.5	76.10±0.3	75.26±0.2	**87.63±0.1**
	40%	57.92±1.4	74.09±0.7	59.15±0.8	58.25±0.2	58.30±0.3	**84.68±0.6**
	60%	36.54±0.62	60.57±0.6	50.54±0.7	42.51±1.2	37.21±0.9	**80.92±0.8**
CIFAR-10	0%	90.39±0.6	**90.27±0.0**	89.03±1.2	89.06±0.9	89.46±0.6	89.41±0.2
	20%	73.12±1.3	84.61±0.3	79.41±0.1	79.19±0.4	82.21±0.4	**85.13±0.6**
	40%	65.07±3.3	81.84±0.1	74.69±1.3	76.67±0.8	75.81±0.3	**83.36±0.5**
	60%	52.55±1.6	72.41±0.7	40.42±0.4	70.57±0.3	68.32±0.6	**72.84±0.6**
CIFAR-100	0%	68.20±0.4	68.54±0.1	68.48±0.2	68.31±0.2	67.89±0.2	**68.60±0.3**
	20%	52.88±0.5	60.25±0.2	58.74±0.3	58.49±0.4	57.32±1.1	**62.20±0.5**
	40%	42.85±0.3	51.27±0.3	45.42±0.6	46.44±0.7	45.77±1.1	**53.01±0.7**
	60%	30.09±0.2	44.22±0.7	34.49±1.1	42.65±0.9	40.29±1.2	**45.21±0.4**

Contact: Xingjun (Daniel) Ma

Email : xingjun.ma@unimelb.edu.au