Additional experimental results

In this document, we will present here the empirical results regarding the CIFAR-10 dataset, layer-wise ablation study, random data, and random labels. We have also submitted our source code package as supplementary material to secure the reproducibility. Our code, trained models, and collected data will be released publicly.

1. Experiments of MLPs on CIFAR-10

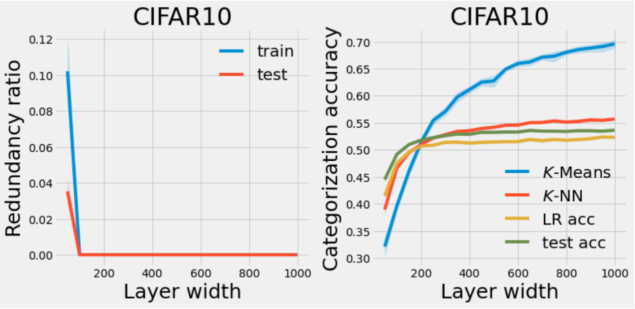
Implementation details

- Model: MLP with five hidden layers.
- Optimizer: Adam.
- Learning rate schedule: initial learning rate = 0.01, decayed by 10 times after every 20 epochs.
- Regularization:
  - Batch Normalization.
  - Weight decay with hyperparameter 1e-6.

Summaries of results

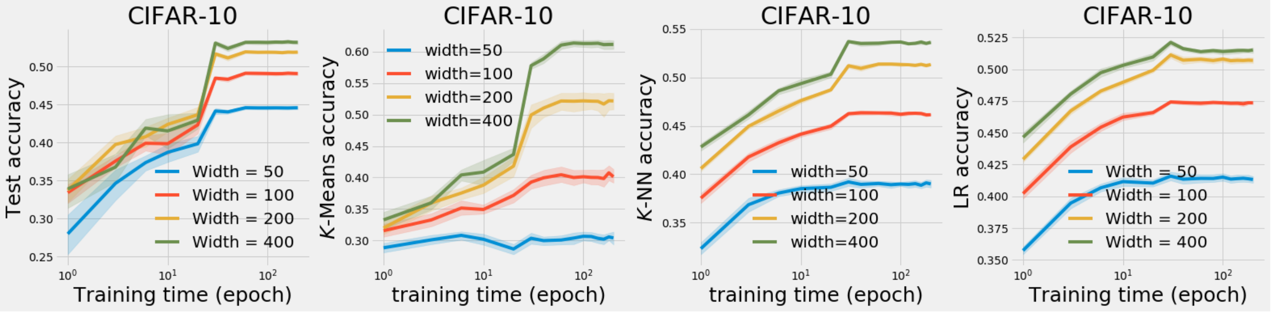
Our experiments show that (1) the redundancy ratio can be almost 0 on CIFAR-10; (2) the categorization accuracy on neural code is almost the same the test accuracy of the corresponding trained model on the raw data; and (3) the two encoding properties have significant correlations with model size, training time, and training sample size.

Influence of layer width



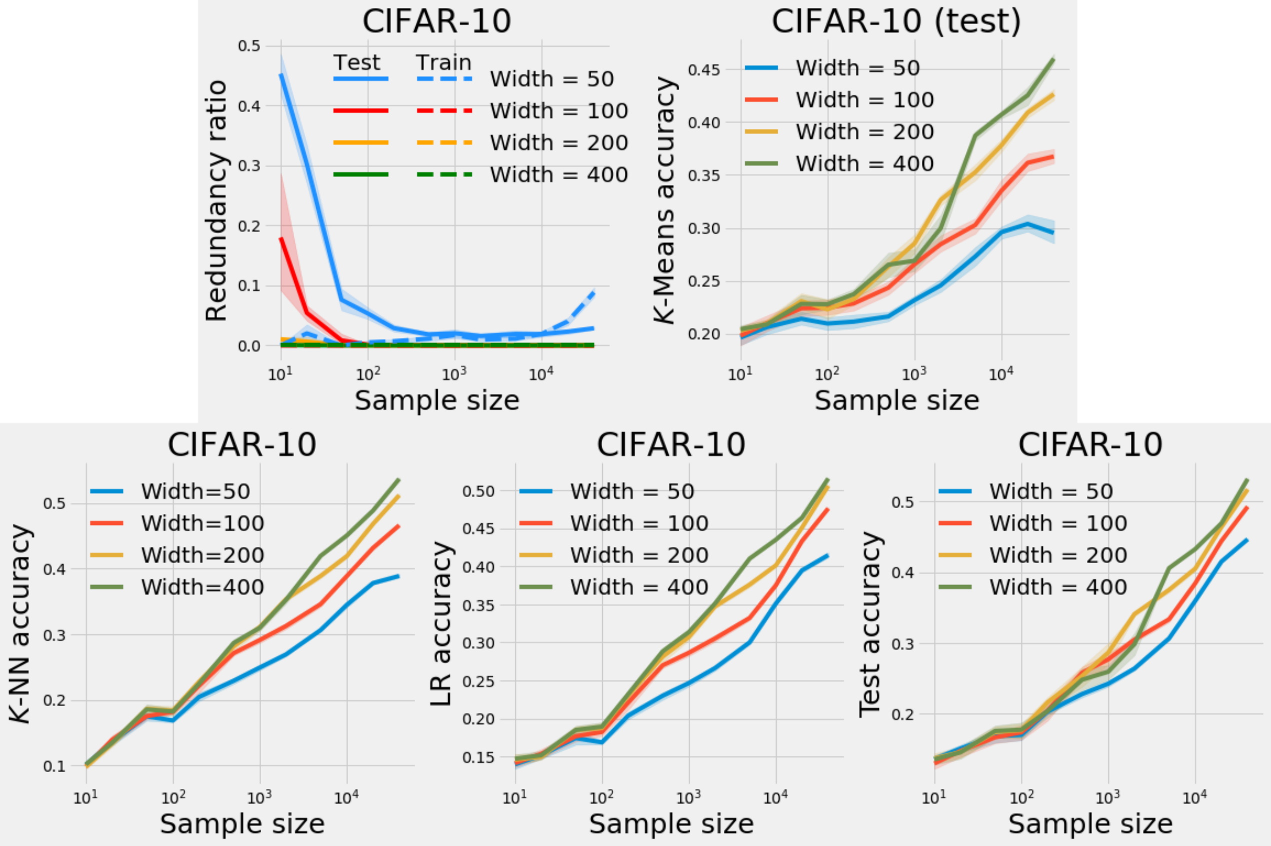
Influence of training time

The layer width is set as 80. The redundancy ratio is 0 in all cases.



Influence of sample size

The layer width is set as 80. The redundancy ratio is 0 in all cases.



2. Layer-wise ablation study

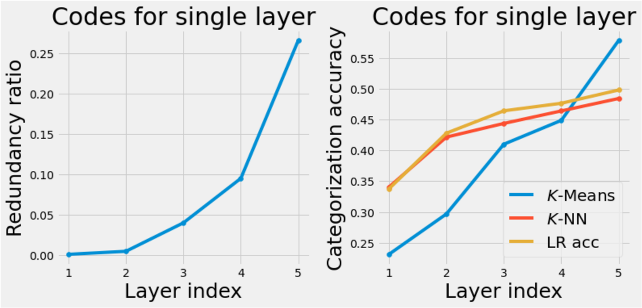
The implementation is the same as Section 1, expect the layer width is set as 40 to ensure the redundancy ratio is not always 0.

Summaries of results

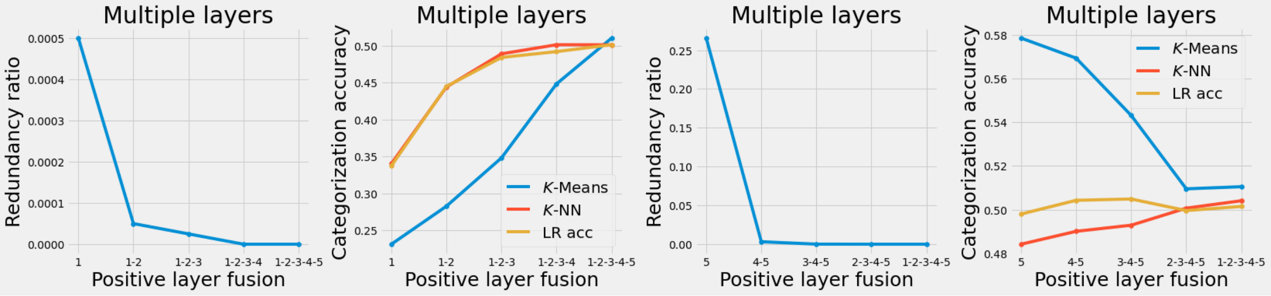
Our experiments show that (1) the earlier layers have fairly low redundancy ratios but relatively poor categorization accuracies; (2) higher layers have relatively poor redundancy ratios though the categorization accuracies are fairly good (approximately equal to the categorization accuracies of the neural code formed by all layers); (3) from the first/last layer towards the full network, the two encoding properties gradually change; (4) the correlations between the encoding properties and model size, training time, and training sample size still hold in the neural code of one single layer; and (5) one can hardly observe both encoding properties are satisfied in the neural code formed by a part of the neural network.

The second property also suggests that one can use the categorization accuracies of final layers to estimate the categorization accuracies of the neural code formed by all layers.

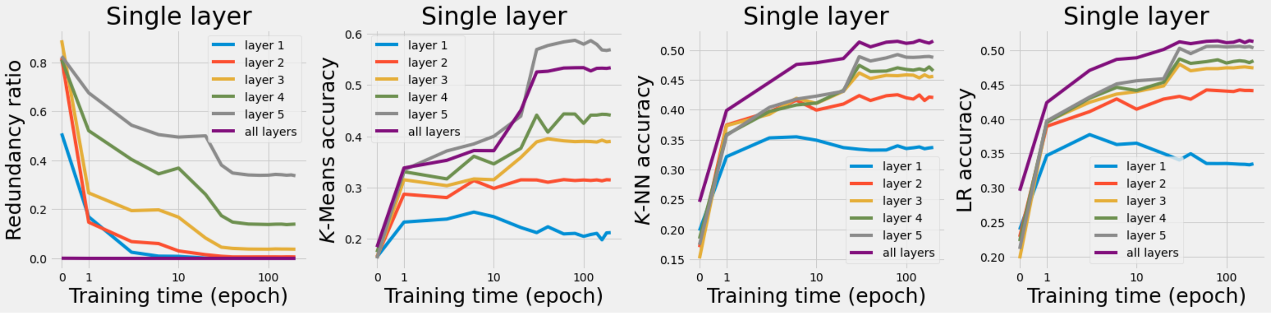
Encoding properties of one single layer



Encoding properties of multiple layers

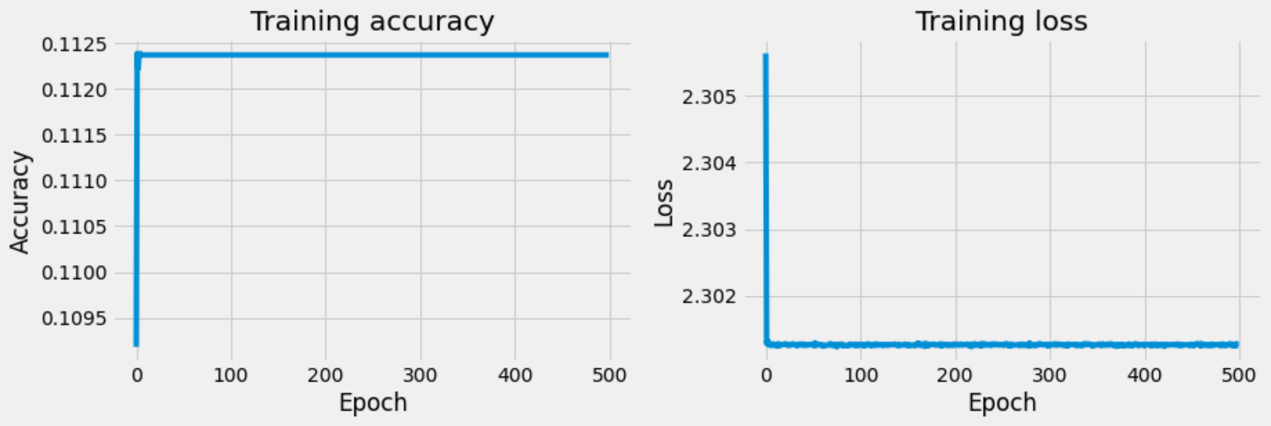


Influence of training time on the encoding properties of one single layer



3. Experiments on random data

We generated images whose pixels are drawn from the uniform distribution $U(0,1)$ with dimension of 784 and sample size of 60,000. We tried to train many MLPs and CNNs on the generated data. The training fails to converge in every case. We show an example of MLP as follows.



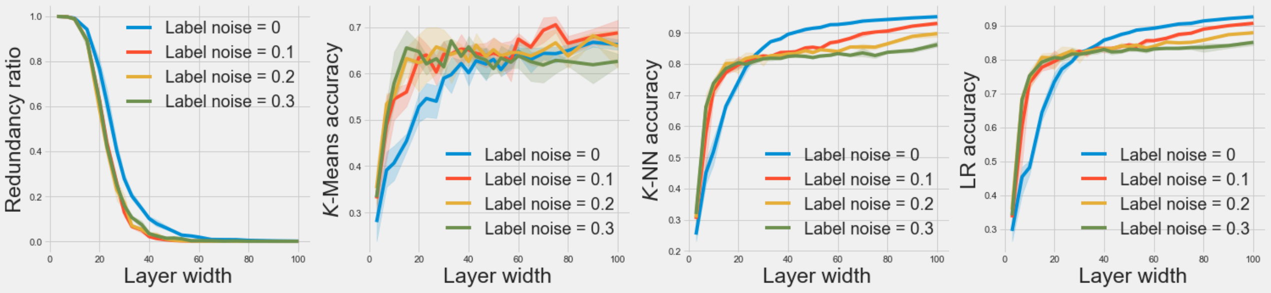
4. Experiments on noisy labels

We introduce label noise into MNIST with different noise rates. Then, one-hidden-layer MLPs with layer width from 3 to 100 are trained on the generated data.

Summary of results

Our experiments show that the encoding properties still stand though become relatively worse. It suggests that the structure of the input-data can drive the organization of the hashed space.

Encoding properties and the influence of layer wise


Redundancy ratio
Label noise = 0
Label noise = 0.1
Label noise = 0.2
Label noise = 0.3
Layer width
K-Means accuracy
Label noise = 0
Label noise = 0.1
Label noise = 0.2
Label noise = 0.3
Layer width
K-NN accuracy
Label noise = 0
Label noise = 0.1
Label noise = 0.2
Label noise = 0.3
Layer width
LR accuracy
Label noise = 0
Label noise = 0.1
Label noise = 0.2
Label noise = 0.3
Layer width