Understanding deep learning requires re-thinking generalization.

Consider handwritten digit recognition

1] Assign random label while training.

2] Training accuracy = 100%

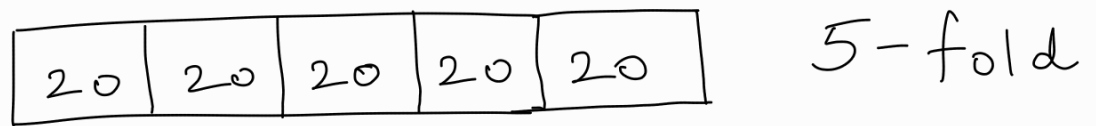The NN memorized which 7 to call 1 and which 7 to call 2, so on.

This implies that training accuracy means nothing. NN can brute force memorize.

3] NN which don't learn cannot generalize.

Doing Experiments

1]

Train	val	Test
80	10	10

2] K-fold validation

20	20	20	20	20

5-fold

Train on 4 parts, test on 5th
Repeat for all parts.
Removes bias which may arise
due to single test set

Consider handwriting recognition. We have
data of 200 people. We now have 2

choices to split data

1) Split with random data from
random people (80/20/20)

2] Split user wise. Use 160 people data for training, 20 for validation, 20 for testing.

→ We should select ② as the purpose of the model is to generalize. In 2nd option, the model's generalization will be tested as it will have unseen data in validation & test set.