Non-linear regression

Approach:
- data: 18k samples, 5k(size) one-hot features, integer target.
- data randomly split into train/test sets
- train/test data got normalized based on avg/std value of train data
- applied augmentations of mixup type
- neural network has the following structure:
    - linear layer (size, size//2), relu ,drop-out
    - linear layer (size//2, size//4), relu ,drop-out
    - linear layer (size//4,1)
- neural network is trained
    - on random mini batch of 512 input elements
    - optimizer – Adam
    - learning rate = 1e-3
    - loss – MSE