NN training with weight decay and cosine annealing