

learning rate

$1e-3$
 $8e-4$
 $6e-4$
 $4e-4$
 $2e-4$
 $0e+0$

0

2M

4M

6M

8M

10M

12M

number of samples trained

Decay Type

fixed

exponential

cyclic

