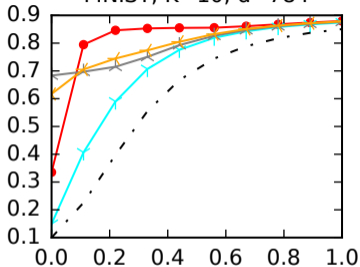
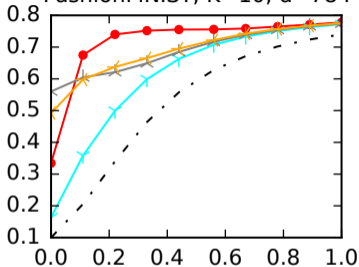


reward of the learned policy

MNIST, K=10, d=784



FashionMNIST, K=10, d=784



inverse-temperature parameter η_0

inverse-temperature parameter η_0



Ours



Sakhi et al. 1



Logging



London et al.



Sakhi et al. 2