

training loss

10^{-1}

- Softmax $d = 2$
- △ Hardmax $d = 2$
- Softmax $d = 4$
- Hardmax $d = 4$

epochs

