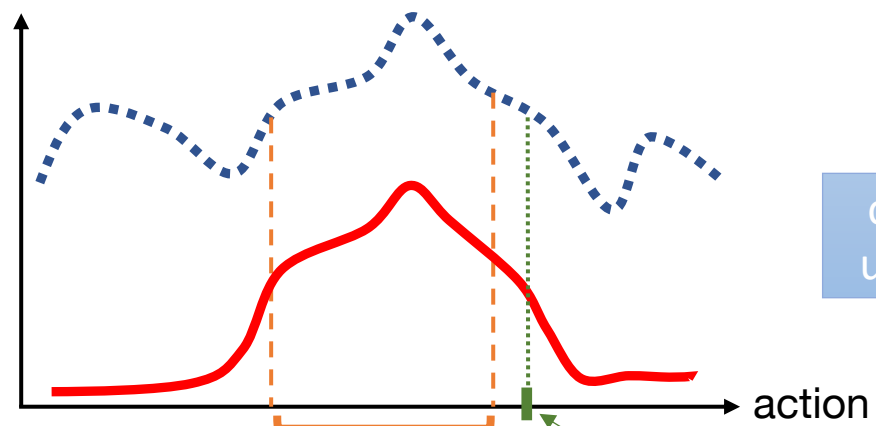


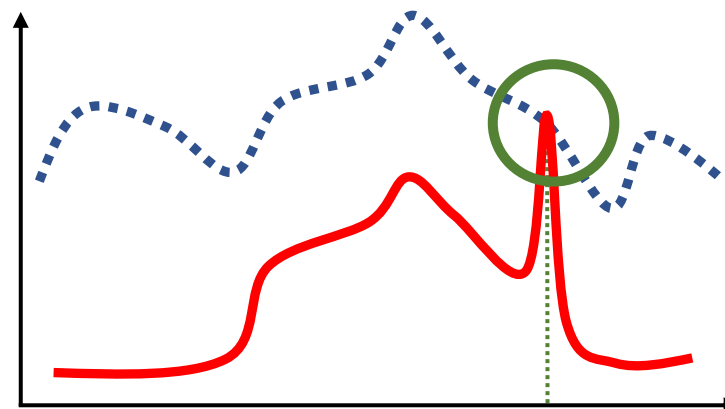
CQL

Pre-trained Q-function



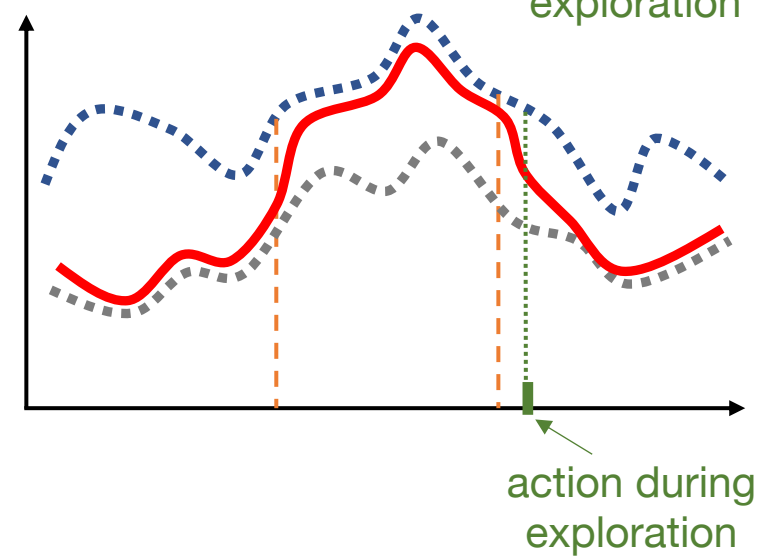
online
update

Online fine-tuned Q-function

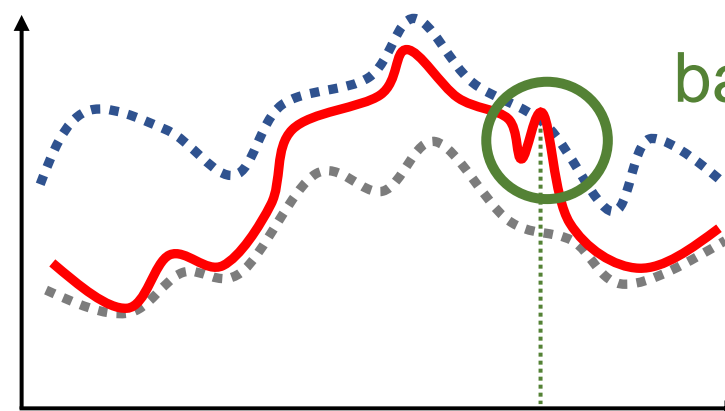


bad action appears optimal
under learned Q-function

Cal-QL



online
update



bad action *doesn't* appear optimal
under learned Q-function

..... ground truth reward — learned Q-function reference Q-function