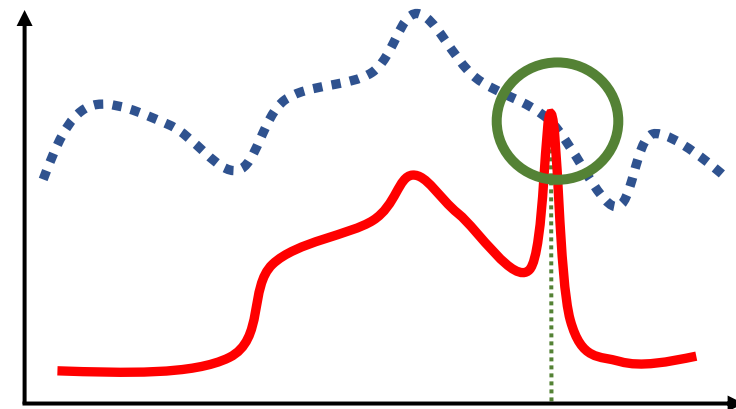
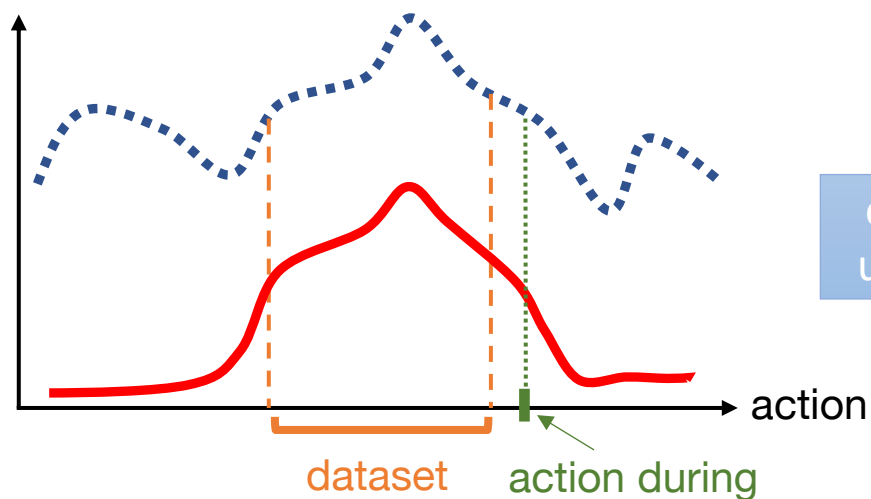


Pre-trained Q-function

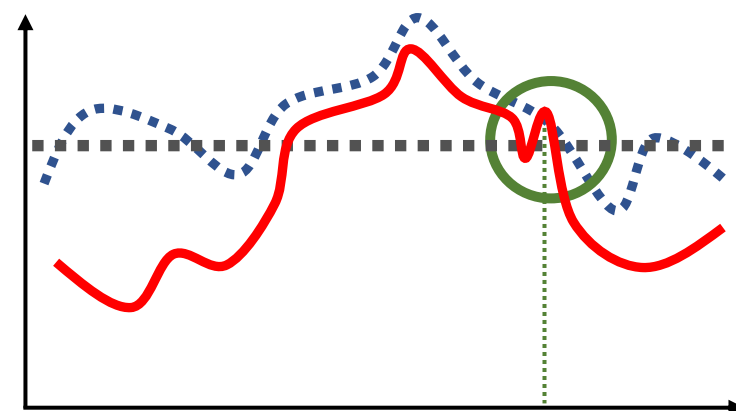
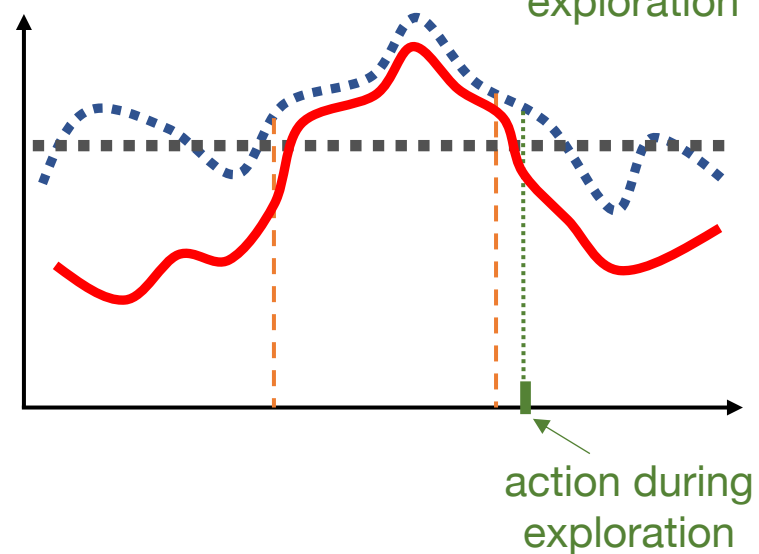
Online fine-tuned Q-function

CQL



bad action appears
optimal
under learned
Q-function

Cal-QL



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

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