GridWorld environment, M = 20

$\epsilon_p = 0$
$\epsilon_p = 0.01$
$\epsilon_p = 0.316$

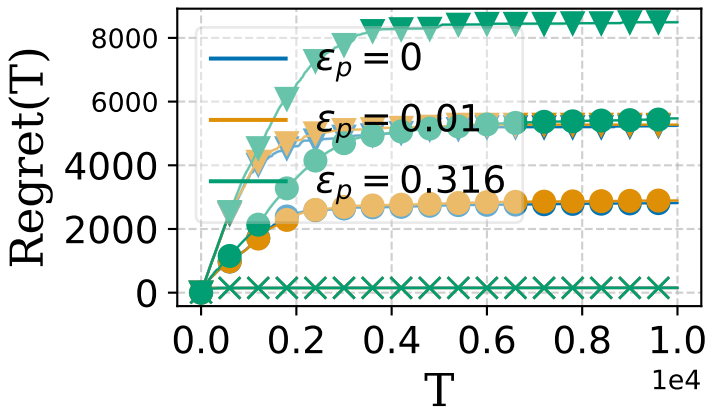Regret(T)

T

1e4

Common regret (lower is better) for M = 20 agents as a function of T for different hetereogeneity levels: crosses represent Fed-UCBVI, circles represent FedQ-Bernstein and triangles represent FedQ-Advantage.