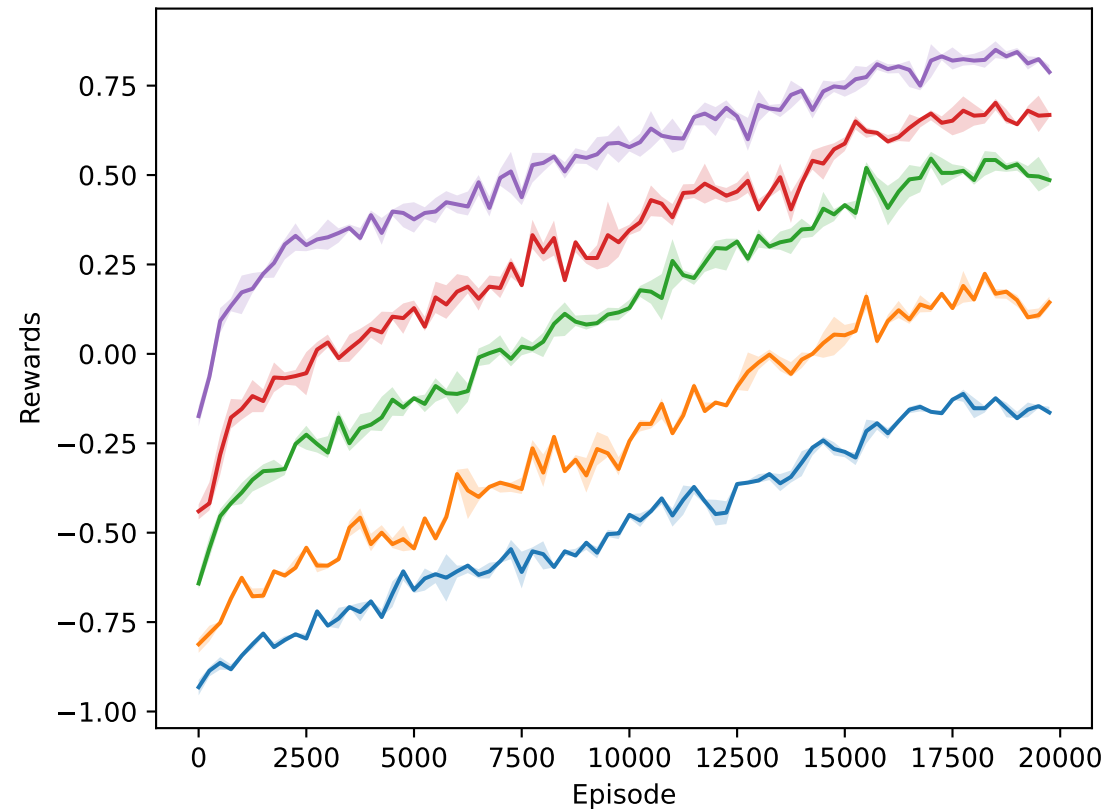
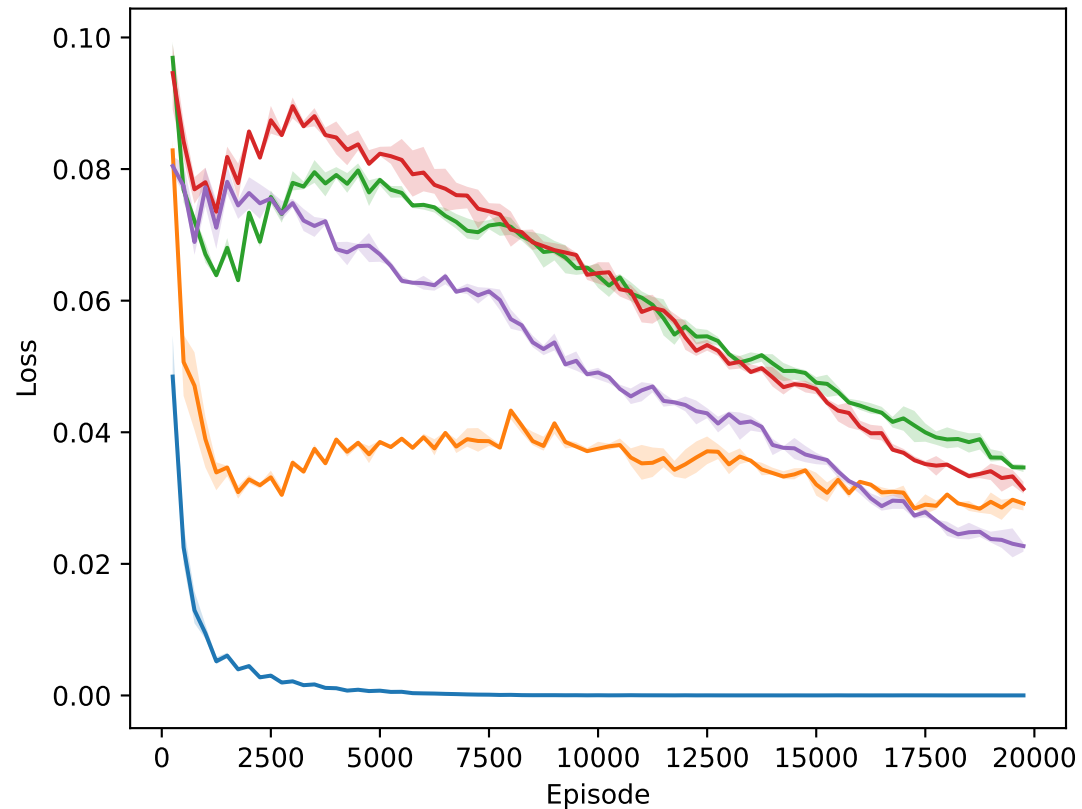


Average rewards during training



Average loss during training



$\epsilon_{opt} = 0.0$ $\epsilon_{opt} = 0.2$ $\epsilon_{opt} = 0.5$ $\epsilon_{opt} = 0.7$ $\epsilon_{opt} = 1.0$