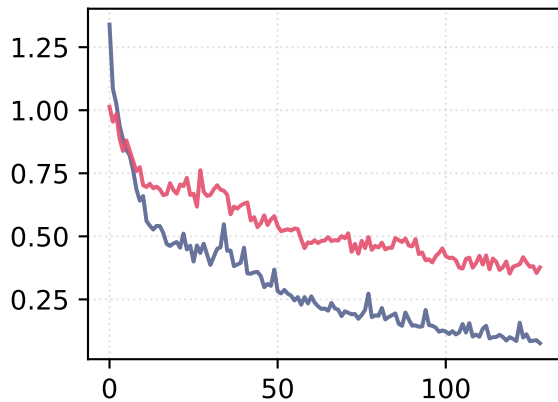
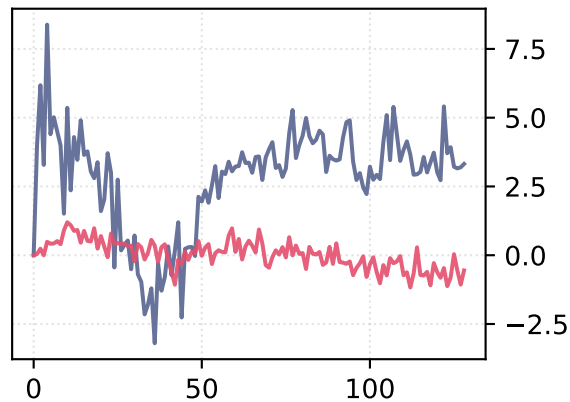




Training loss



Kullback-Leibler divergence



Reward mean | polynomial regression, deg = 2

