Safe risk MDP Q-value, property R2, shallow 3-layered network 2.5 2.0 1.5 1.0 One-hot encoding 0.5 Ordinal encoding One-hot encoding (Q-table) Ordinal encoding (Q-table) 0.0 Modest solution = 1.8100 200 300 400 500 600

Number of episodes

Average reward