**returns**: η sweep on LightDark training 15 returns 10 = 0.01= 0.0001= 0.0000110 20 30 40 policy iterations