odes	Rate=3e-5 Ra	ite=3e-4 Rate=	1.5e-3 Ra	ate=3e-3 La	ayers=2 L	_ayers=3	Layers=4	Lambda=0.9 La	ambda=0.95 La	ambda=0.99	
10000	5.415194	7.140235	5.499813	8.995246	7.140235	8.23896	10.25772381	8.958658	12.62222	8.23896	
20000	7.635468	15.25099	7.505676	8.675706	15.25099	19.18152	26.77825165	21.6792	32.74821	19.18152	
30000	10.49198	23.64768	2.953012	8.510129	23.64768	27.4387	37.06264496	37.80915	44.28489	27.4387	Training Rewards for Different Learning Rates
40000	8.868241	38.34888	1.634561	14.28962	38.34888	36.20419	42.81414413	46.48121	49.47441	36.20419	■ Rate=3e-5 ■ Rate=3e-4 ■ Rate=1.5e-3 ■ Rate=3e-3
50000	11.88365	35.50576	5.144563	17.30124	35.50576	45.4353	48.72386169	51.18036	51.79501	45.4353	
60000	15.94528	43.11034	7.127337	7.342403	43.11034	45.41676	49.10373688	54.6779	56.28595	45.41676	80
70000	18.73909	42.75176	8.158346	0.840596	42.75176	48.26775	43.38427734	56.6665	49.95293	48.26775	
80000	20.76465	51.92688	6.451464	3.931207	51.92688	61.06288	46.04838562	53.8875	47.06948	61.06288	ē 60 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
90000	26.49182	51.28976	7.32783	6.320073	51.28976	60.24873	47.51274109	60.44774	50.78257	60.24873	
100000	30.42443	54.6176	7.299128	12.07667	54.6176	58.34872	49.23637772	67.88873	53.01249	58.34872	g /
110000	28.77578	54.28605	7.690099	15.60126	54.28605	59.03178	46.99996567	61.06589	50.74318	59.03178	® 40 √
20000	32.02405	50.04247	10.78063	20.86796	50.04247	63.76246	50.19797897	62.91433	47.23584	63.76246	© 20 /
130000	33.0963	52.91704	12.18995	17.83196	52.91704	68.14897	54.33443451	66.90531	49.21726	68.14897	
40000	34.67532	56.07186	12.90877	24.43511	56.07186	68.42681	59.20777512	63.08569	54.97041	68.42681	
150000	36.95332	56.28983	9.610162	28.14876	56.28983	76.166	61.80220413	64.49875	63.33727	76.166	100000 200000 300000 400000 500000
.60000	36.00465	60.01566	16.53867	24.10799	60.01566	75.44068	62.95045853	58.63895	65.44718	75.44068	
70000	37.26893	58.53596	15.31613	33.04401	58.53596	68.60542	60.99375534	55.78897	63.62641	68.60542	Episodes
.80000	38.51305	57.6837	16.83614	29.61319	57.6837	67.03314	53.92502594	55.24907	61.45224	67.03314	
90000	41.90938	63.45747	13.23755	28.58937	63.45747	72.68812	59.80508804	63.23795	69.82854	72.68812	T
00000	46.25872	62.70743	17.52044	25.06708	62.70743	67.93919	66.68214417	61.45573	73.91547	67.93919	Training Rewards For Different Numbers of 128-Node Hidden
10000	45.94804	61.74963	20.54564	9.301348	61.74963	69.51923	64.76459503	53.46328	64.32893	69.51923	Layers
20000	49.14951	59.4769	22.03242	14.7644	59.4769	71.24894	57.36940002	63.43726	72.09914	71.24894	■ Layers=2 ■ Layers=3 ■ Layers=4
30000	46.97144	61.54991	23.11149	13.52618	61.54991	66.47061	51.96737289	58.94074	67.85333	66.47061	80
40000	47.1156	60.3587	27.00704	13.82785	60.3587	69.6069	53.19114685	63.1626	69.8016	69.6069	80
250000	46.29868	63.38519	18.24375	16.75842	63.38519	70.52962	54.66718674	67.80852	67.46711	70.52962	
260000	46.11984	69.07545	19.33484	20.06456	69.07545	72.26617	49.57540131	65.05578	71.08263	72.26617	M 60 60
270000	48.50027	62.82215	23.72989	25.50038	62.82215	66.26192	50.0597229	66.94254	73.91426	66.26192	a si
80000	49.74507	67.78049	23.15968	28.30386	67.78049	70.49046	53.00688171	70.80002	71.41906	70.49046	te 40 //
90000	51.07119	67.3196	22.85515	30.78831	67.3196	69.71986	64.38413239	67.19362	77.56838	69.71986	m No.
00000	47.83974	65.44593	26.74899	29.40678	65.44593	70.33672	65.3383255	68.73339	75.67281	70.33672	8, 20
10000	53.41156	61.17725	25.67274	29.47687	61.17725	74.03214	72.31935883	66.76691	71.02903	74.03214	E P
20000	56.57856	64.33185	28.82597		64.33185	71.8642	65.64149475	63.22902	67.51113	71.8642	₹ 0
30000	52.35275	62.77294	30.01167		62.77294	73.04367	62.06480789	64.60882	69.15475	73.04367	100000 200000 300000 400000
40000	52.31053	62.91048	29.92782		62.91048	75.80083	65.50892639	62.62731	68.21137	75.80083	Episodes
50000	55.75488	58.58732	32.86681		58.58732	70.98306	69.13980103	65.36726	68.60232	70.98306	Epitodos
60000	56.04295	68.37631	34.07088		68.37631	73.45922	75.32252502	62.76199	71.55531	73.45922	
70000	58.55051	63.03429	29.94916		63.03429	73.77725	75.28151703	64.73438	66.49715	73.77725	
80000	59.24367		32.97557			72.79428	65.44547272	69.99806	71.94234	72.79428	Training Rewards For Different Lambda Values
90000	59.91391		31.57952			73.89142	72.14099121	69.58847	69.88371	73.89142	 Lambda=0.9 Lambda=0.95
00000	55.75359		33.27816			77.41689	71.48846436	68.20274	68.18182	77.41689	80
10000	51.42982		30.04012			76.05053	72.06130981	68.14977	69.59795	76.05053	00
20000	59.28112		31.45793			76.82988	73.637146	70.52507	74.35906	76.82988	
30000	62.87943		30.34667			79.951	67.73345947	69.80583	74.79109	79.951	₩ 60 × 60 × 60 × 60 × 60 × 60 × 60 × 60
40000	60.66518		29.49758			77.08186	70.00673676	71.02065	75.74386	77.08186	å V
50000	61.77407		28.56534			79.16452	70.76842499	69.38807	73.17287	79.16452	\$ 40 M
60000	60.76419		32.19407			77.52731	75.83774567	72.1227	73.41914	77.52731	e " ///
70000	58.62282		31.19484			79.46449	69.53227234	71.62997	74.09499	79.46449	Original Control of the Control of t
180000	55.89629		28.58682			78.05402	71.9826889	72.6183	76.48145	78.05402	© 20 /
190000	58.91031		34.63874			78.05402	66.8267746	72.6183	74.2317	78.05402	
	60.89978		34.63874			77.48757		71.13165	76.78321	77.48757	
500000	60.89978		32.4952			/9.829/5	75.13246918	/1.13105	/6./8321	/9.829/5	100000 200000 300000 400000 500000
											Episodes