

Deep learning 1-3 Report

컴퓨터소프트웨어학부 209016735 현수빈

1. Parameter 값

- Initial function Parameter W, b : 모두 -1,1 사이의 랜덤 float값으로 초기화, 아래 표 참고
- Estimated unknown function parameter W, b: 아래 표 참고
- Empirically determined best hyper parameter: learning rate 값의 초기에는 0.01로 시작하지만 학습 중 train_accuracy의 값이 90%보다 낮을 경우 learning rate를 $(100 - \text{accuracy}) / 100$ 값으로 올리고(1보다 작지만 정확도가 낮을수록 크다), 90%보다 높을 경우 잘 학습된 것으로 생각하여 0.01로 낮추었다.
- Accuracy: 아래 표 참고

2. 결과

Table 1	M=10, n=1000, k= 5000	M=100, n=1000, k= 5000	M=10000, n=1000, k= 5000
Accuracy (with m train samples)	90.00%	97.00%	99.24%
Accuracy (with n test samples)	57.70%	96.00%	98.60%
Initial W	[-0.2327263, -0.10259526, -0.44384937, 0.15886201, -0.83077766, 0.98060883]	[-0.58916367, 0.92454935, 0.35049549, 0.74560418, 0.39592179 -0.91620742]	[0.5203127, -0.32880533, 0.7383955, -0.53781028, 0.44431584 0.3091663]
Initial B	[-4.96034142, 3.59007011, 6.785055309810648]	[0.1814745, 0.43186276, 0.35090729719119995]	[0.8208172, 0.8371929067776485, -0.82906464]
Estimated W1	[[-0.05665315 -1.36573492] [-1.00569718 -0.24132544]]	[[-0.81328621 -0.00450749] [1.22903811 -0.03746342]]	[[-6.62290033e+00, 8.05949270e-05] [-5.38227981e+00, 9.39332816e-04]]
Estimated W2	[[-2.85510045 2.34405325]]	[[-4.06036814 -5.18623334]]	[[5.74321495][4.97106764]]
Estimated B1	[[-0.46608525][-1.39949267]]	[[3.63931327][5.70502808]]	[[-5.73773077 -5.68156437]]
Estimated B2	[[-1.11897693]]	[[6.41004019]]	[[8.14835428]]
Final alpha	0.01	0.01	0.01

첫번째는 test 시 일반적으로 training accuracy와 test accuracy의 편차가 높게 나오는 경향이 있었다. 두번째의 경우 training sample이 10배 늘었기에 훨씬 더 적은 편차가 나왔다. 세번째가 제일 training set 수가 크기에 low bias, low variance로 가장 좋은 결과가 나왔다.

Table 2	M=10000, n=1000, k= 10	M=10000, n=1000, k= 100	M=10000, n=1000, k= 5000
Accuracy (with m train samples)	66.64%	88.51%	98.76%
Accuracy (with n test samples)	68.10%	89.00%	98.50%
Initial W	[-0.66696593, -0.41340708, -0.73380505, 0.50205675, 0.19632685, 0.105975]	[0.1557328 -0.05082701, 0.45481423 -0.37586246, -0.06416742 0.02972485]	[-0.81953075, 0.5985005, -0.71267076 0.87285075, -0.73231381 -0.03586973]
Initial B	[0.14447115, 0.08102007, -0.8236347022978212]	[0.5510413, -0.85714738, 0.10036360712941206]	[0.69060392, 0.9072775, -0.17098997691024298]
Estimated W1	[[-0.64921247 -0.43472712] [-0.72475519 0.51202377]]	[[0.40343307 -0.00693602] [0.55037156 -0.09705055]]	[[-1.10026572, 3.22031143e-04] [-0.776519746, -8.38238498e-04]]
Estimated W2	[[0.32019343 0.23540152]]	[[-0.45264864 0.45355284]]	[[-5.4612371 5.27799873]]
Estimated B1	[[0.13241021] [0.07490521]]	[[0.69752445] [-0.96120662]]	[[5.10759231] [-3.39928265]]
Estimated B2	[[-0.44143611]]	[[0.06708464]]	[[2.37533714]]
Final alpha	0.3313000000000001	0.11599999999999994	0.01

첫번째의 경우, Iteration 수가 매우 작기에 train accuracy와 test accuracy가 전반적으로 낮았다. 두번째의 경우 10배로 iteration 수를 늘렸지만 살짝 좋아졌으나 큰 효과는 없었다. iteration이 높은 3번째의 경우 정확도가 가장 높은 모습을 볼 수 있었다.