

Deep learning 1-3 Report

컴퓨터소프트웨어학부 209016735 현수빈

1. Parameter 값

- Initial function Parameter W, b : 모두 -10,10 사이의 랜덤 float값으로 초기화, 아래 표 참고
- Estimated unknown function parameter W, b: 아래 표 참고
- Empirically determined best hyper parameter: learning rate 값의 경우 0.01로만 하면 정확도가 낮게 나와 다른 방식으로 조절했다. 초기에는 0.01로 시작하지만 학습 중 train_accuracy의 값이 90%보다 낮을 경우 learning rate를 3으로 높이고 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%	100.00%	99.99%
Accuracy (with n test samples)	52.20%	97.70%	100.00%
Initial W	[-4.22794131, -2.25486191, -1.10160845, -7.14932857, 7.31559415, 1.38357432]	[-5.32926593, -0.20064492, 5.39428761, -0.49318284, -5.96003076, 8.43325334]	[8.09501059, 0.71166117, -9.28053406, 6.92730286, 4.01072246, -8.1428672]
Initial B	[-4.96034142, 3.59007011, 6.785055309810648]	[-4.56624497, -7.28468074, -6.120801240959944]	[-6.88005033, -8.94641542, 4.560953153462119]
Estimated W1	[[-1.18442033 5.73359666] [9.75442478 -3.7419483]]	[[-2.47375396 -0.03528262] [3.19547101, 0.01547767]]	[[-6.62290033e+00, 8.05949270e-05] [-5.38227981e+00, 9.39332816e-04]]
Estimated W2	[[-3.71103365, 10.23170237]]	[15.78265662, 16.46006305]]	[8.80100635, -8.07246393]]
Estimated B1	[[-9.29385601][6.86913606]]	[[-11.53014912][-16.54561984]]	[[-32.97407686][26.60400409]]
Estimated B2	[4.41319237]]	[[-8.03954022]]	[3.45156271]]
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)	68.51%	75.78%	99.87%
Accuracy (with n test samples)	70.90%	77.60%	99.90%
Initial W	[5.22783001, 6.57516177], 3.67896043 0.90740627, -0.90869999 3.81591751]	[7.99734208 1.61395457] [-1.15826578 1.44827383, 1.35279059 -3.96649022]	[8.9644607, 2.71987007, -9.69900623, -3.82159503, 9.92804973 1.53089734]]
Initial B	[-0.06159148, -7.646564, 2.0928919101167143]	[2.36680725, 6.18265577, 3.6121427343217842]	[-5.66931896, 9.84372889, -9.164150325414758]
Estimated W1	[5.74787358, 6.11905926] [1.86288358, 0.20652]]	[7.56869799 1.59217167] [-2.02620478 0.31775201]]	[1.88002400, 5.63564822e-03] [-2.83834682, -6.22138415e-04]]
Estimated W2	[[-1.73268607 2.97043428]]	[[-3.00451598 -5.07789273]]	[[-5.93342624 -6.4166406]]
Estimated B1	[[-0.05943024] [-8.34746415]]	[3.30946471] [9.08439683]]	[9.10953083] [14.10692409]]
Estimated B2	[[-0.01681546]]	[5.11307745]]	[9.00147307]]
Final alpha	3	3	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.