

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: alpha값의 경우 0.01로만 하면 정확도가 낮게 나와 다른 방식으로 조절했다. 초기에는 0.01로 시작하지만 80% 이상 training하게 되면 alpha값을 0.01로 줄였다. 학습 중 training accuracy 출력 값(training 수의 1/10번마다 출력)이 이전 training accuracy보다 같거나 크면 맞는 방향이므로 1을 더해 가속도 있게 트레이닝 한다.
- 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)	100.00%	100.00%	99.99%
Accuracy (with n test samples)	69.30%	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	[[-1.87455741 -0.26753269] [-0.64315423, -6.94172173]]	[[-2.47375396 -0.03528262] [3.19547101, 0.01547767]]	[[-6.62290033e+00, 8.05949270e-05] [-5.38227981e+00, 9.39332816e-04]]
Estimated W2	[[16.48732074, -1.82989679]]	[[15.78265662, 16.46006305]]	[[8.80100635, -8.07246393]]
Estimated B1	[[10.11941465][4.32471508]]	[[-11.53014912][-16.54561984]]	[[-32.97407686][26.60400409]]
Estimated B2	[[-7.0088074]]	[[-8.03954022]]	[[-3.45156271]]
alpha	1.01	1.01	1.01

첫번째는 test 시 일반적으로 training accuracy가 test accuracy보다 약 30% 높게 나오는 경향이 있었다. 두번째의 경우 training sample이 10배 늘었기에 훨씬 더 좋은 결과가 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)	68.83%	96.72%	100.00%
Accuracy (with n test samples)	69.60%	96.70%	100.00%
Initial W	[2.20537081, 4.0198248, -0.45458782, 1.73950428, -7.74475888, -8.32509218]	[-5.06131326 -2.02277463, -2.09600714, -2.51082001, 3.28553198 -2.37507382]	[8.9644607, 2.71987007, -9.69900623, -3.82159503, 9.92804973 1.53089734]]
Initial B	[5.19562726, 3.80880783, 6.36075245218143]	[-3.53125686, 2.94963703, 9.694785103510299]	[-5.66931896, 9.84372889, -9.164150325414758]
Estimated W1	[[4.29044415, 2.10912156] [-2.97579636, 0.77503716]]	[[-1.56484842e+00, -1.12882410e-01] [-1.67291921e+00 7.70469121e-04]]	[[6.09834010e+00, -1.80457488e-03] [-5.14125421e+00, -4.26853893e-03]]
Estimated W2	[[-6.22212842, -6.85854975]]	[[4.91596365, -4.87055834]]	[[9.3445607, 8.59811022]]
Estimated B1	[[5.0840918] [4.00765267]	[[-6.20013879] [7.67290151]]	[[-30.44162101] [-25.48758896]]
Estimated B2	[[7.49962902]]	[[1.66826512]]	[[-4.7901887]]
alpha	1.01	1.01	1.01

첫번째의 경우, Iteration 수가 10으로 매우 작기에 high bias, high variance의 모습을 보이고 있다. Iteration 수를 늘릴 수록 점점 정확도가 높아지고 제일 iteration이 높은 3번째의 경우 정확도가 가장 높은 모습을 볼 수 있었다.