

Deep Learning

Practice#1 Report

2016024766

김서현

1. Time Comparison (m=1000, k=2000)

Element-wise version	Vectorized version
6.0567s	0.1152s

Vectorized version이 Element-wise version보다 약 52.5배 빠르게 실행되는 것을 알 수 있습니다.

2. Estimated unknown function parameters W & b

$W_1 = 5.04$, $W_2 = 5.04$, $b = -0.18$

3. Empirically determined best hyper parameter, α

α 가 0.01보다 작으면 accuracy가 95% 이하로 낮아지고, α 가 10보다 크면 sigmoid function을 계산할 때 overflow가 나는 경우가 발생했습니다. 따라서 overflow가 나지 않는 범위 내에서 accuracy가 99% 이상이 나오도록 **iteration 횟수가 많을 때는 $\alpha=0.5$, 적을 때는 $\alpha=5$** 로 설정하는 것이 적합하다고 생각합니다. (ex. iteration 횟수가 500 미만일 때는 $\alpha=5$, 500 이상일 때는 $\alpha=0.5$)

4. Accuracy ($\alpha = 0.5$ 로 설정) (각 경우마다 5번 실행 후 accuracy의 평균을 계산해 기입)

	M=10, n=100, k=2000	M=100, n=100, k=2000	M=1000, n=100, k=2000
Accuracy (with 'm' train set)	100.0%	100.0%	99.9%
Accuracy (with 'n' test samples)	90.6%	99.8%	99.8%
	M=1000, n=100, k=20	M=1000, n=100, k=200	M=1000, n=100, k=2000
Accuracy (with 'm' train set)	90.2%	98.1%	99.9%
Accuracy (with 'n' test samples)	89.0%	97.2%	99.8%

5. Discussion

- Accuracy 비교를 통해 training set의 개수 보다는 iteration 횟수가 accuracy를 높이는 데에 있어 더 중요한 요소라는 결과를 도출할 수 있었습니다.
- α 값이 너무 작으면 주어진 iteration 횟수 안에 최적의 w, b 를 찾지 못하고, α 가 너무 크면 sigmoid 함수를 계산할 때에 overflow가 발생한다는 것을 알 수 있었습니다.
- $K=20$ 인 경우에는 α 가 5일 때의 accuracy가 α 가 0.5일 때의 accuracy보다 훨씬 크다는 것을 확인했습니다.
- Gradient descent를 하는 과정 내내 loss function (cost)의 값은 점점 작아지고 accuracy는 점점 증가한다는 것을 확인하였습니다.
- Vectorized 하여 계산하는 것이 Element-wise 방식으로 계산하는 것보다 훨씬 빠르다는 것을 체감하였습니다.