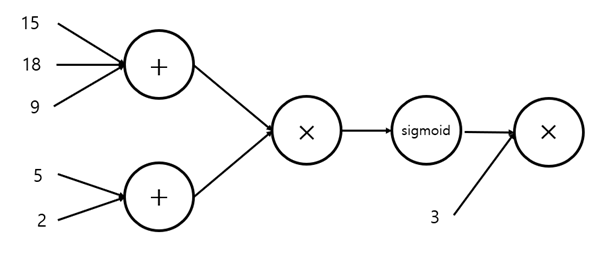
**2019년 DAIA 3기 과제**

**<이론>**

1. 다음의 계산 그래프의 Backpropagation을 계산하세요.  
   

|  |  |
| --- | --- |
| 답 |  |
|  | Sigmoid 함수( )의 로컬 그래디언트는 이다. |

1. Gradient Vanishing에 대하여 자세히 서술하고, 이를 방지할 수 있는 기법을 모두 적고 그에 대한 설명도 자세히 서술하세요.

|  |  |
| --- | --- |
| 답 | Gradient Vanishing이란 activation function을 선택하는 문제에서 발생한다. 예를 들어 tanh함수와 함께 Sigmoid 함수는 모든 입력 값을 [0,1] 또는 [-1,1]로 압축하여 매핑한다. 그 결과 매우 넓은 input space의 값이 극도로 작은 범위로 매핑되어 버린다.  결과적으로 input space에서는 큰 변화가 있더라도 output에는 작은 변화를 보이게 된다.  따라서 이러한 문제를 해결하기 위해 “squashing”의 특징을 갖지 않는 activation function을 사용해야한다. 최대값이 제한되어 있지 않은 함수를 찾으면 된다. 가장 흔하게 쓰이는 함수로는 ReLU가 있다. 0보다 작을 때는 0을 사용하고, 0보다 큰 값에는 해당하는 값을 그대로 사용하는 방식이다.  다른 기법으로는 weight를 초기화 시키는 방법에 대한 연구이다. 구체적인 방법으로는 RBM(Restricted Boatman Machine)과 Xavier/He initializion이 있다.  RBM이란 weight 값을 조절하는 방법으로 현재 layer와 다음 layer에 대해서만 동작한다. (forward) 현재 layer에 들어온 x값에 대해 weight을 계산한 값을 다음 layer에 전달한다. (backward) 전달받은 값을 weight 값을 계산해서 거꾸로 이전(현재) layer에 전달한다. forward와 backward 계산을 반복해서 진행하면, 최초 전달된 x와 예측한 값(x hat)의 차이가 최소가 되는 weight을 발견하게 된다. 이렇게 만들어진 weight 값을 초기값으로 설정한다. 이러한 과정은 Learning이 아닌 Fine Tuning이라 불린다.    Xavier/He initializion 방식은 He’s initialization라고도 불린다. 입력 값을 fan-in, 출력 값을 fan-out이라고 부른다. Xavier 는 입력 값과 출력 값 사이의 난수를 선택해서 입력 값의 제곱근으로 나누고, He는 입력 값을 반으로 나눈 제곱근을 사용한다. 분모가 작아지기 때문에 Xavier보다 넓은 범위의 난수 생성한다. 이는 입력의 개수와 출력의 개수에 따라 적절한 초기 값을 지원해주는 방법이다. weight의 값을 random으로 주긴 하지만 매우 적절한 초기 값을 주는 도구이다. |

1. Activation function(Sigmoid, Tanh, ReLU, Leaky ReLU)에 대해 자세히 서술하고 각 function의 장, 단점을 서술하세요.

|  |  |
| --- | --- |
| 답 | https://cdn-images-1.medium.com/max/1440/1*nrxtwp6rzqdFhgYh0x-eVw.png  Activation function의 그래프 형태는 위와 같다.  1. Sigmoid 함수  - 설명  대표적인 Logistic 함수로서 이는 Binary classification 에 적절한 함수다. 일정 값을 기준으로 0인지 1인지 구분함으로써 분류하는 방식이다. 특정 임계치를 넘을 때만 활성화되기 때문에 활성 함수 중 하나로 구분되는 함수다. 경사 하강법(Gradient Descent Algorithm) 계산 또는 역전파(Backpropagarion) 계산 과정에서 Sigmoid 함수의 미분이 사용된다.  - 장단점  모든 실수 입력 값을 0보다 크고 1보다 작은 미분 가능한 수로 변환하는 특징을 가진다. 모든 입력 값을 미분 가능한 0~1 사이의 값으로 반환하기 때문에 분류 문제와 비용 함수(Cost Function)에 많이 사용된다. 또한, Sigmoid의 반환 값은 확률 형태이기 때문에 결과를 확률로 해석할 때 유용하다.  https://taewanmerepo.github.io/2017/09/sigmoid/differential_sigmoid.jpg  Sigmoid 함수의 미분 그래프  Sigmoid의 미분 그래프의 미분 계수의 최댓값을 보면 0.25 이다. Backpropagation을 계산하는 과정에서 활성함수의 미분 값을 곱하는 과정이 있다. 이때, 은닉층의 깊이가 깊다면 오차율 계산이 어렵다는 문제가 발생한다. chain rule을 이용해 계속 값을 곱할 때 결과 값이 0에 수렴하기 때문이다. 즉 입력 값이 최종적인 계층에 미치는 영향이 적어지는 Gradient vanishing problem 이 발생한다.  2. Tanh 함수  Hyperbolic Tangent(tanh) 함수는 확장된 Sigmoid 함수로 Sigmoid 함수 대신 사용할 수 있다. tanh와 Sigmoid의 차이점은 Sigmoid의 출력 범위가 0에서 1 사이인 반면 tanh와 출력 범위는 -1에서 1사이라는 점이다. 즉 함수 값이 zero-centerd 되어 있다는 점에서 다르다. 때문에 tanh 비선형함수는 sigmoid 보다 선호도가 높다  https://taewanmerepo.github.io/2017/12/tanh/020.jpg  - 장단점  Sigmoid와 비교하여 tanh와는 출력 범위가 더 넓고 경사면이 큰 범위가 더 크다. tanh 그래프의 미분계수를 보면 최댓값은 1이다. Sigmoid 미분계수와 비교하여 최대값이 4배가 크다.  Sigmoid 보다는 양호하지만 tanh를 활성 함수로 사용하는 경우 은닉층의 깊이가 깊다면, tanh 역시 Gradient vanishing problem 이 발생한다.  3. ReLU 함수  이 함수는 쉽게 말해 0보다 작은 값이 나온 경우 0을 반환하고, 0보다 큰 값이 나온 경우 그 값을 그대로 반환하는 함수이다. 0보다 큰 값일 경우 1을 반환하는 sigmoid와 다르다. 따라서 내부 hidden layer에는 ReLU를 적용하고, 마지막 output layer에서만 sigmoid 함수를 적용하면 이전에 비해 정확도가 훨씬 올라가게 된다.  - 장단점  선형 함수이므로 미분 계산이 편리하여 계산 복잡성이 낮아 시간 단축이 된다. X가 양수이면 Gradient가 1로 일정하므로 Gradient vanishing 현상이 없다.  ReLU는 음수들을 모두 0으로 처리하기 때문에 한번 음수가 나오면 더 이상 그 노드는 학습되지 않는다. 이러한 점을 보완한 것이 Leaky ReLU가 있다.  4. Leaky ReLU 함수  http://i.imgur.com/SXq4jmJ.png  ReLU 함수에서 0보다 작은 값이 입력으로 들어왔을 때, 함수값이 0으로 정해져 뉴런의 값이 더 이상 업데이트 되지 않는다. 이러한 문제점을 해결하기 위해 제시된 함수인 Leaky ReLU 함수는 0.01로 작은 기울기를 부여한다. |

1. MLE, MAP에 대하여 자세히 서술하고, 그 차이점에 대해서도 서술하세요.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 답 | - MLE는 Maximum Likelihood Estimation의 약자로 Likelihood는 이미 주어진 표본적 증거를 바탕으로 모집단에 대해 어떠한 통계적 추정이 그럴듯한 정도를 말한다. 즉, 어떤 가설을 전제했을 때 그 전제하에서 주어진 증거가 얼마나 타당한 가에 대한 정도이다.  P(H)에 대한 확률을 모르는데, 어떠한 데이터가 주어진 경우, 이 데이터를 통해서 확률 P(H)를 추정하는 과정을 Maximum Likelihood Estimation 이라 한다.  예를 들면, 동전 던지기에 대해 생각해보자. 앞 또는 뒤가 나올 확률은 같다고 가정하여 0.5라고 생각한다. 만약 100번의 동전던지기를 수행할 때, 앞면이 56번 나왔다면 동전을 던졌을 때 앞면이 나올 확률은 얼마라고 할 수 있는가? 이를 MLE를 이용해 풀면, L(T|E)에서 증거 E는 앞면이 56번 나왔다는 사실이고, 이론 T를 변화시키면서 어느 이론이 가장 그 확률이 높은 지 찾는 과정이다. L(T|E)는 P(E|T)를 최대화하는 과정이기 때문에 이론 T가 “P(H) = 0.5”일 때를 계산하면 다음과 같다.  다른 이론 T에 대해 값을 구해보면 다음 표와 같다. 아래 표에서 가장 높은 L을 가지는 이론 T는 P=0.56일 때이다. 즉, 동전의 앞면이 나올 확률이 0.5라는 것을 전혀 모르는 상황에서 이러한 증거가 있을 때 “동전을 던져서 앞면이 나올 확률은 0.56이다.”라고 말할 수 있다.   |  |  | | --- | --- | | p | L | | 0.48 | 0.0222 | | 0.50 | 0.0389 | | 0.52 | 0.0581 | | 0.54 | 0.0739 | | 0.56 | 0.0801 | | 0.58 | 0.0738 | | 0.60 | 0.0576 | | 0.62 | 0.0378 |   위 예제의 출처는 <http://statgen.iop.kcl.ac.uk/bgim/mle/sslike_3.html> 이다.  하지만, MLE는 observation에 따라 값이 너무 민감하게 변한다는 단점이 있다. 극단적으로 n번 던져서 앞면이 n번 나오는 경우, MLE는 앞면만 나오는 동전으로 판단해버린다.  - MLE의 단점을 해결하기 위해 나온 MAP은 Maximum A Posterior의 약자이다. MAP는 주어진 데이터에 대해 최대 확률을 가지는 것을 찾는다. parameter estimation을 하면 오직 주어진 데이터만을 잘 설명하는 parameter값을 찾게 된다. |

1. Loss function(MSE, Cross-Entropy(KL-Divergence)에 대하여 자세하게 서술하세요.

|  |  |
| --- | --- |
| 답 | 신경망 학습에서는 최적의 매개변수를 탐색할 때 손실함수의 값을 가능한 작게 하는 매개변수의 값을 찾는다. 손실 함수란 신경망이 학습할 수 있도록 해주는 지표이다. 모델의 출력 값과 사용자가 원하는 출력 값과의 차이, 즉 오차를 말한다. 이 손실 함수의 값이 최소화되도록 하는 가중치와 편향을 찾는 것이 바로 학습이다. 보통 loss 함수, cost 함수라고 불린다.    - MSE는 Mean Squared Error의 약자로 평균 제곱 오차를 나타낸다.  ( y: 신경망 출력 값, t: 정답 레이블, k: 데이터 차원의 수)  평균 제곱 오차 함수는 회귀에서 많이 쓰이는 손실함수이다. 거리 차이를 제곱하면 어느 부분에서 오차가 두드러지는지 확실히 알 수 있다.  - Cross-Entropy(KL-Divergence)는 교차 엔트로피 오차를 나타낸다.  ( log: 밑이 e인 자연로그, y:신경망 출력 값, t: 정답 레이블)  교차 엔트로피 오차 함수는 분류에서 많이 쓰이는 손실함수이다. 분류(Classification) 문제에서 one-hot encoding 했을 경우에만 사용할 수 있는 오차 계산법이다. |

**<실습>**

https://www.kaggle.com/c/titanic(케글의 타이타닉문제)링크를 들어가, 제시된 문제를 해결해보세요.

1. 제시된 데이터를 바탕으로 해당 인원이 살아남을지 죽을지 예측하는 문제
2. 코드에 주석은 무조건 달아주세요. 인터넷의 코드 참고 가능합니다.
3. 위 링크로 들어간 후 'data'탭 에서 data다운로드 가능

내부 팀별 리더보드를 통해서 1등 팀에게는 스타벅스 기프티콘 증정  
<외부코드를 참고는 가능하지만 단순히 가져 다만 써서 정확도를 뽑아낸 경우는 과제로 인정 불가>

<과제 제출>

마감 기한 : ~2.28

이론 과제 : 워드파일로 번호별로 작성하여 슬랙에 게시

실습 과제 : 분석 코드, 발표자료(ppt) 게시

발표 자료의 경우 3.1에 팀마다 발표를 할 예정이며, 팀마다 어떻게 분석을 진행했고 분석기법이나 딥러닝 모델 등 구체적으로 어떻게 본인들의 결과를 얻게 되었는지 등을 준비하시면 될 겁니다.

(분석 공모전 대회 발표자료 준비한다고 생각하시면 편할 것임)

궁금한 사항은 단톡방에 질문해주시면 알려드리겠습니다~