

## 1. 개요

2장에서는 펄스의 관측 문제로 참, 거짓을 구분하는 이진 판단 문제였습니다. 3장에서는 철판이 불량인가 판단하는 이진 판단이 아닌 어떤 불량 여부인지도 가려내야 하는 선택분류 문제입니다. 이진 판단처럼 신경망이 직접 후보 항목의 번호를 지정하거나, 각 후보 항목의 확률 값을 계산하기는 어렵습니다. 선택 분류 신경망은 이진 판단처럼 각 후보에 대한 로그 척도의 상대적 추천 강도, 즉 로짓값을 추정하도록 구성됩니다.

## 2. 정보량과 엔트로피

### 2.1 엔트로피(entropy)

정보량의 과학적 개념은 '어떤 것의 가능한 대안의 수 또는 모든 경우의 수의 총량'을 의미합니다. 어떤 사건의 경우의 수가 많다면 특정한 사건의 발생 가능성이 희박하다는 이야기이므로 정보량은 곧 확률론적 속성을 가지고 있다고 볼 수 있습니다. 따라서 정보량은 확률함수로 나타내는 것이 자연스럽습니다.

$$h(x) = -\log p(x) \quad \text{식(1)}$$

여기서  $p(x)$ 는  $x$ 의 확률밀도함수입니다.  $p(x)$ 가 0에 가까워질수록, 즉 사건의 빈도수가 작을수록 정보량은 많아지고  $p(x)$  값이 커진다는 것은 사건의 빈도수가 커지는 것이고 정보량은 작아지게 됩니다. 엔트로피는 정보량의 기대 값으로 정의하며 수식은 식(2)와 같습니다.

$$H(p) = E_{x \sim p(x)}[-\log p(x)] = - \int_x p(x) \log p(x) dx \quad \text{식(2)}$$

### 2.2 교차 엔트로피(cross entropy)

확률밀도함수  $q(x)$ 의 정보량을 확률밀도함수  $p(x)$ 의 관점에서 기대값으로 표현한 것을  $p(x)$ 와  $q(x)$ 의 교차 엔트로피라고 하며 다음 식(3)과 같이 정의됩니다.

$$H(p, q) = E_{x \sim p(x)}[-\log q(x)] = - \int_x p(x) \log q(x) dx \quad \text{식(3)}$$

기계학습에서는 교차 엔트로피를 실제 정답의 확률만을 고려한 손실(비용)함수입니다. 이때  $p$ 는 true probability로써 true label에 대한 분포를,  $q$ 는 현재 예측모델의 추정값에 대한 분포를 나타내게 됩니다.

### 3. 소프트맥스함수(SoftmaxFunction)

이진 분류를 위해 사용한 시그모이드 함수 대신 다중 분류에 주로 사용하는 활성화 함수인 소프트맥스함수를 사용하게 됩니다. 소프트맥스 함수는 로짓값 벡터를 확률 분포 벡터로 변환해주는 비선형 함수로 시그모이드함수에서 유도되었습니다. 출력값이 여러 개로 주어지고 목표치가 다범주인 경우 각 범주에 속할 사후 확률을 제공하며 출력값의 총 합은 1이 됩니다.

$$y = \frac{e^{x_n}}{\sum_{k=1}^N e^{x_k}} \quad \text{식4)}$$

식4)를 다시 풀어보면 k번일확률/전체확률임을 알 수 있습니다.

### 4. 실험

4.1은 교재 3장의 코드를 사용했으며 4.2는 소프트맥스 함수대신 2장의 시그모이드 함수를 사용했습니다. 손실함수는 교차엔트로피 대신 1장의 MSE를 활용했습니다. 실험 결과는 다음과 같습니다.

#### 4.1 Softmax Function + Cross Entropy

```
Epoch 1: loss=15.984, accuracy=0.306/0.320
Epoch 2: loss=15.509, accuracy=0.326/0.197
Epoch 3: loss=15.984, accuracy=0.306/0.348
Epoch 4: loss=15.004, accuracy=0.348/0.197
Epoch 5: loss=15.286, accuracy=0.336/0.202
Epoch 6: loss=15.390, accuracy=0.332/0.440
Epoch 7: loss=15.509, accuracy=0.326/0.442
Epoch 8: loss=15.628, accuracy=0.321/0.455
Epoch 9: loss=15.360, accuracy=0.333/0.322
Epoch 10: loss=15.316, accuracy=0.335/0.455
Final Test: final accuracy = 0.455
```

#### 4.2 Sigmoid Function + Mean Squared Error

```
Epoch 1: loss= nan, accuracy=0.103/0.090
Epoch 2: loss= nan, accuracy=0.079/0.090
Epoch 3: loss= nan, accuracy=0.079/0.090
Epoch 4: loss= nan, accuracy=0.079/0.090
Epoch 5: loss= nan, accuracy=0.079/0.090
Epoch 6: loss= nan, accuracy=0.079/0.090
Epoch 7: loss= nan, accuracy=0.079/0.090
Epoch 8: loss= nan, accuracy=0.079/0.090
Epoch 9: loss= nan, accuracy=0.079/0.090
Epoch 10: loss= nan, accuracy=0.079/0.090
Final Test: final accuracy = 0.090
RuntimeWarning: overflow encountered
```

## 5. 결론

MSE는 출력결과와 데이터 차이 제곱의 평균으로 정답과 오답의 모든 확률을 고려하게 되고, 교차 엔트로피는 실제 정답의 확률만을 고려하게 됩니다. 여러 개중 하나를 고르는 다중 선택문제에서 (4.2)는 overflow에러를 출력하며 손실함수 값을 전혀 찾지 못했으며, 매우 낮은 정답률을 보이고 있습니다. 이는 시그모이드함수에 손실함수로 MSE를 사용하게 되면 편미분해서 기울기가 0인 지점으로 경사 하강을 시도할 때 지역국소점에 걸리게 되어 최적의 값을 찾지 못하게 됩니다. 또한 saddle point에 빠지는 위험이 큰데 이 뜻은 극소점은 아니지만 기울기가 굉장히 완만한 지점을 만났을 때 경사하강법이 잘 작동하지 않아 학습이 굉장히 느려지게 됩니다. 결론적으로 분류문제를 만나게 되었을 때는 손실함수를 교차엔트로피를 사용하는 것이 옳다는 것을 알 수 있습니다.

4.1의 소프트맥스 함수와 교차엔트로피를 사용한 결과도 높은 정확도를 보이지는 못합니다. 27가지 요소는 27차원의 벡터공간이며, 이를 7가지로 적절한 분할을 한다는 것은 많은 정보를 요구하게 됩니다. 이번 실험에 사용한 1941개의 데이터수는 너무 적어 충분한 학습이 이루어지기 어려웠으며, 단층퍼셉트론으로는 구조가 단순하다는 한계를 보였습니다. 마지막으로 실험에 있어 상황에 맞는 활성화 함수를 선택하는 걸로 끝나지 않으며 적합한 손실함수를 사용하는 것 역시 중요한 요소임을 알게 되었습니다.