



Softmax

Kim Jeongseob

작성 2020.11.14

요약 :

N개의 데이터가 있을 경우, 각 값을 지수적으로 증가시켜 편차를 늘린 다음 normalization하는 함수. 즉, 주어진 값들의 편차를 증가시켜 큰 값은 상대적으로 더 크게, 작은 값은 상대적으로 더 작게 만들어서 가장 큰 값을 더욱 더 두드러지게 만드는 효과를 주게 한다.

$$\text{softmax}_n(x) = \frac{e^{x_n}}{\sum_{k=1}^N e^{x_k}}, n \in \{1, 2, 3, \dots, N\}$$

Softmax가 의미하는 바는 N개의 원소로 이루어진 값의 집합에서 특정 인덱스 n의 원소의 값을 중요도를 구하는 것이라고 볼 수 있다. 예를 들면, N개의 클래스가 있다고 할 때, n번째 클래스에 속할 확률이 가장 높다면, softmax 함수를 거쳐 이 확률을 더욱 더 크게 만든다.

5개의 클래스로 이루어진 경우를 생각해 보자. 신경망에서 softmax를 거치기 전에 계산된 출력이 다음과 같다면,

{1, 2.4, 0.8, 5.6, 0.2}

각각의 경우 softmax 함수를 거치게 되면 다음과 같이 계산된다.

{0.0094511 0.0383262 0.0077379 0.9402381 0.0042467}

4번째 원소의 값이 가장 큰 값이었으므로 softmax 함수를 거친 값도 가장 크다. Classification 문제에 있어서 4번째 클래스에 해당될 확률이 약 94%라는 값으로 계산된다.

다만, softmax에 입력이 되는 값의 크기가 작다면 오히려 편차가 줄어들기도 한다.

{0.1, 0.24, 0.08, 0.56, 0.02} → softmax(x) → {0.17738 0.20403 0.17387 0.28098 0.16374}

기존 값들이 스케일만 달라졌을 뿐인데, 편차가 오히려 줄어들었다.

요약 :

N개의 데이터가 있을 경우, 각 값을 지수적으로 증가시켜 편차를 늘린 다음 normalization하는 함수. 즉, 주어진 값들의 편차를 증가시켜 큰 값은 상대적으로 더 크게, 작은 값은 상대적으로 더 작게 만들어서 가장 큰 값을 더욱 더 두드러지게 만드는 효과를 주게 한다.

$$\text{softmax}_n(x) = \frac{e^{x_n}}{\sum_{k=1}^N e^{x_k}}, n \in \{1, 2, 3, \dots, N\}$$

Softmax가 의미하는 바는 N개의 원소로 이루어진 값의 집합에서 특정 인덱스 n의 원소의 값을 중요도를 구하는 것이라고 볼 수 있다. 예를 들면, N개의 클래스가 있다고 할 때, n번째 클래스에 속할 확률이 가장 높다면, softmax 함수를 거쳐 이 확률을 더욱 더 크게 만든다.

5개의 클래스로 이루어진 경우를 생각해 보자. 신경망에서 softmax를 거치기 전에 계산된 출력이 다음과 같다면,

{1, 2.4, 0.8, 5.6, 0.2}

각각의 경우 softmax 함수를 거치게 되면 다음과 같이 계산된다.

{0.0094511 0.0383262 0.0077379 0.9402381 0.0042467}

4번째 원소의 값이 가장 큰 값이었으므로 softmax 함수를 거친 값도 가장 크다. Classification 문제에 있어서 4번째 클래스에 해당될 확률이 약 94%라는 값으로 계산된다.

다만, softmax에 입력이 되는 값의 크기가 작다면 오히려 편차가 줄어들기도 한다.

{0.1, 0.24, 0.08, 0.56, 0.02} → softmax(x) → {0.17738 0.20403 0.17387 0.28098 0.16374}

기존 값들이 스케일만 달라졌을 뿐인데, 편차가 오히려 줄어들었다.