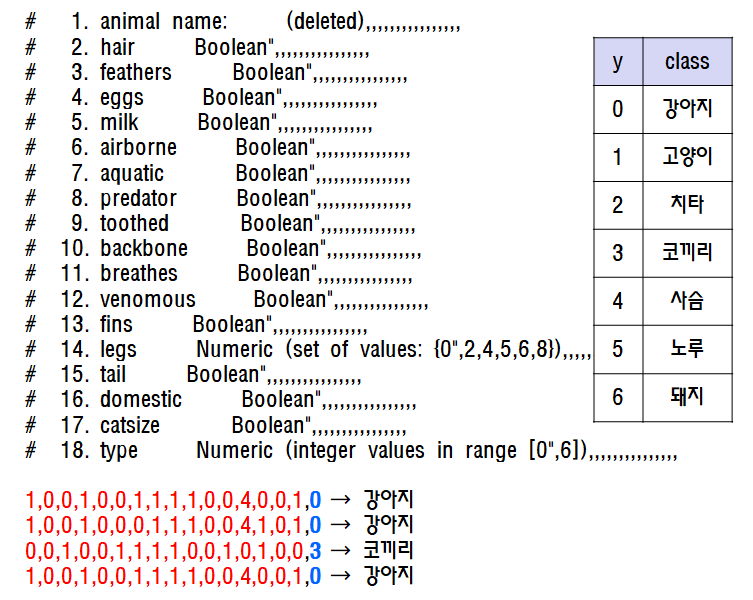
**소프트 맥스**

다음과 같이 7개의 클래스로 구분된 데이터가 있다고 가정합니다.



클래스의 개수가 n인 분류 문제에서, i번째 클래스가 정답일 확률을 Pi로 표현하겠습니다.

softmax 함수의 정의는 다음과 같습니다.

분자) j번째 원소를 의미합니다.

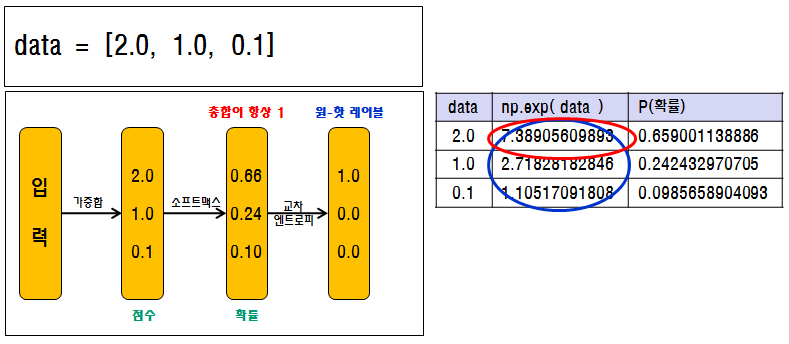
분모) 1부터 n까지의 지수 함수 결과 값의 총합을 의미합니다.

여기서, n=3이라고 가정합니다.

가설 공식은 다음과 같습니다.

여기서, P1, P2, P3 는 각각 i번째 클래스가 정답일 확률을 의미합니다.

따라서, P1+P2+P3 =1이 됩니다.



### 소프트 맥스의 주요 특징

소프트 맥스는 다음과 같은 특징을 가지고 있습니다.

|  |
| --- |
| 주요 특징 |
| 수치/점수 등의 데이터를 확률로 만들어 주는 개념입니다.  모든 확률(probabilities)의 총합은 1입니다.  컴퓨터는 오직 숫자로만 처리가 가능하므로, 그 판단은 사람의 몫입니다. |

One Hot Encoding

n=3이라고 가정하고, 컬럼의 개수를 n개 만큼 만든 다음, 해당 인덱스에만 값 1을 지정하는 방법입니다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 클래스 | A | B | C |  |  |
| 수치 표현 | 0 | 1 | 2 |  |  |
| 인코딩 | 1 0 0 | 0 1 0 | 0 0 1 |  |  |

다중 분류에서는 정답 컬럼의 개수는 1이 아니고, class의 개수와 동일 해야 합니다.

유클리디언 거리를 이용하여 다음과 같이 구해 봅니다.

인코딩 전의 거리 구하기

(A-B)\*\*2 = (0-1)\*\*2 = 1

(A-C)\*\*2 = (0-2)\*\*2 = 4

위의 결과를 보면 A에 대하여 C보다는 B가 더 가깝다는 의미가 됩니다.

다중 분류에서는 거리 개념으로 보았을 때, 동등한 거리이어야 합니다.

이때 필요한 것이 원핫 인코딩입니다.

인코딩 후의 거리 구하기

(A-B)\*\*2 = ((1 0 0) – (0 1 0))\*\*2 = 1 + 1 + 0 = 2

(A-C)\*\*2 = ((1 0 0) – (0 0 1))\*\*2 = 1 + 0 + 1 = 2

(B-C)\*\*2 = ((0 1 0) – (0 0 1))\*\*2 = 0 + 1 + 1 = 2

원핫 인코딩 된 데이터들은 유클리디언 거리 값이 모두 동일합니다.

비용 함수 공식은 다음과 같습니다.