3.5 출력층 설계하기

도입

- Output 을 위한 Node에는 어떠한 Activation Function 이 들어가야 할까?
- 점수로 나타낸다면?
    - softmax? 항등함수? 상관없을까?
- 분류를 위한 확률로 나타낸다면?
    - 관상 App
    - 성격 Test App

회귀

- 입력데이터에 대한 수치를 예측하는 문제
- 값을 예측하기 위함이므로 항등함수를 주로 사용한다.

분류

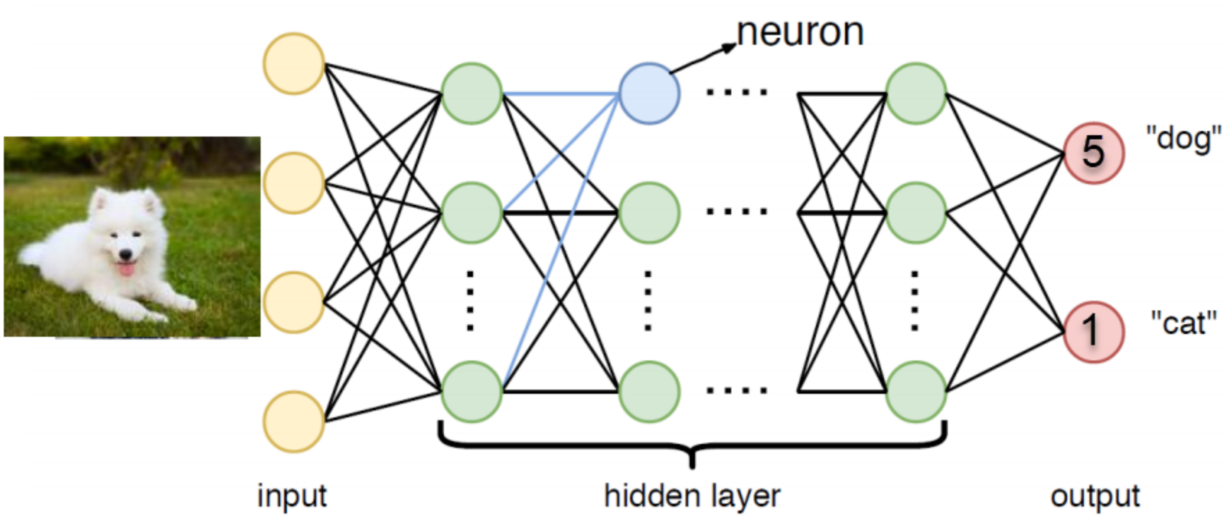
- 입력데이터에 대한 분류
- 어느 Class 에 속할 확률을 나타내기 위함이므로 확률에 관련된 Function 을 사용한다.

분류를 위한 출력함수

- 분류를 위한 가장 좋은 출력값을 어떤 형태이어야 할까?
    - 정수 (-∞, ∞)
    - 확률 (0,1)

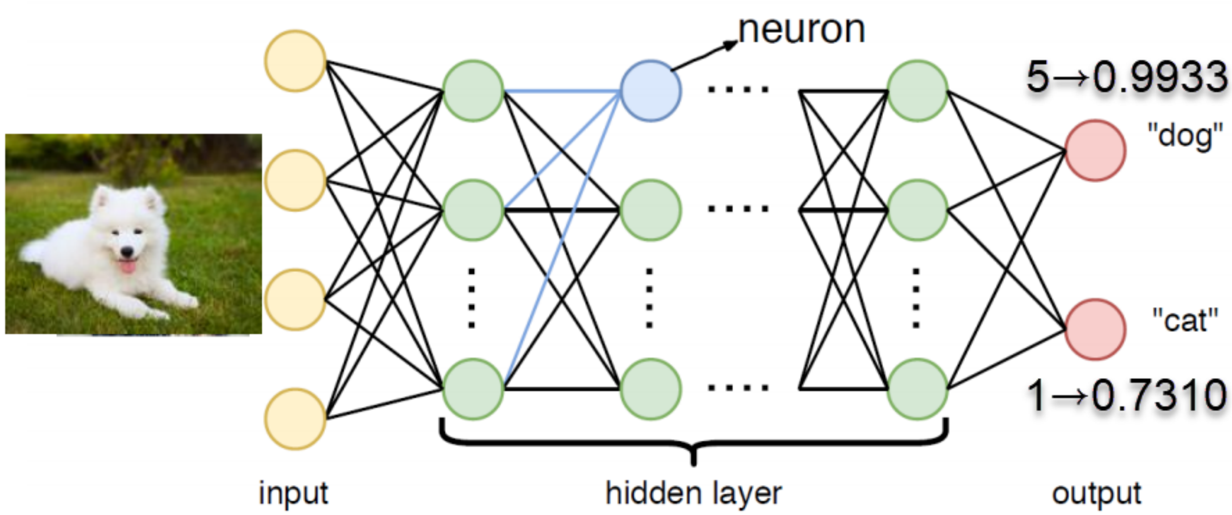
함수에 따른 Output

Score



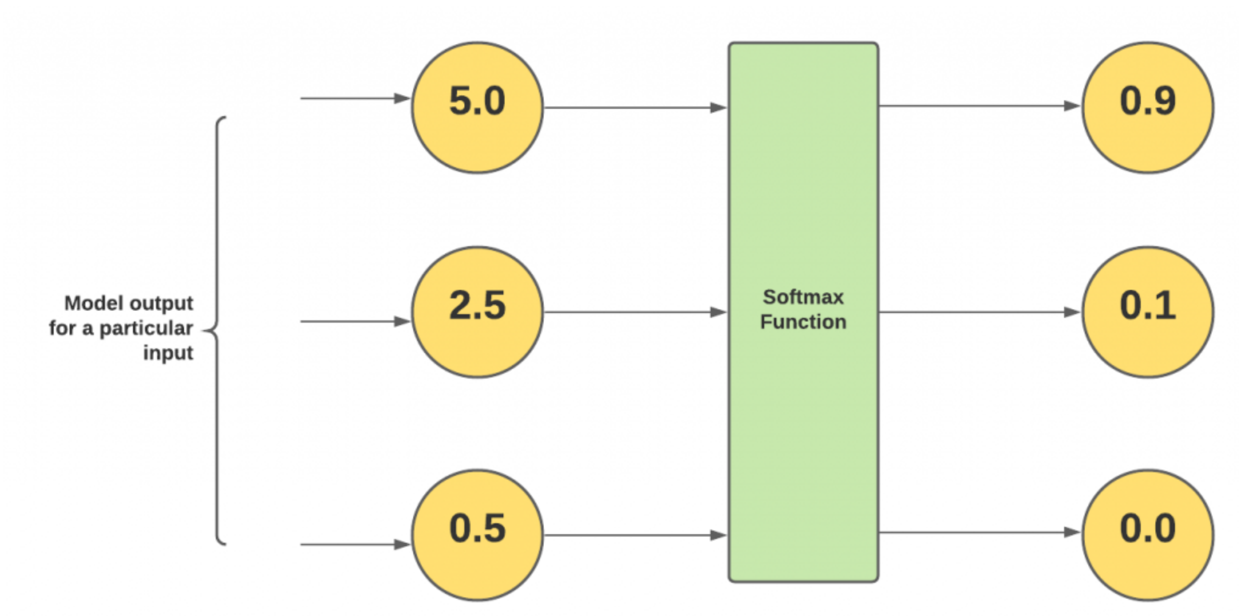
Sigmoid Function

- 확률 x
- 단순히 (0,1) 의 값



Softmax Function

- 확률 o
- 총합이 1 인 값



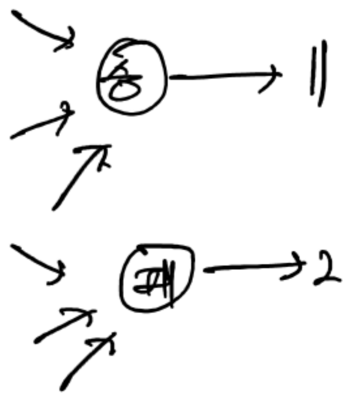
Sigmoid, Logit and Softmaxsoftmax와 sigmoid

Class1 과 Class2 의 분류

- 이진 분류

odd(승산)

- 성공할 확률을 p라 할때, 실패 대비 성공할 비율을 승산 이라고 한다.
- $odd = \frac{p}{1-p}$
- 예
    - Chelsea 의 승리 11, 패배 2 일때 첼시의 패배 확률은?
    - 첼시의 패배 확률 : $\frac{p}{1-p} = \frac{2}{11}$ 이므로 $p = \frac{2}{13}$
    - 첼시의 승리 확률 : $\frac{p}{1-p} = \frac{11}{2}$ 이므로 $p = \frac{11}{13}$



logit Function

- odd에 밑이 e인 log함수를 취한값

$$z = \ln\left(\frac{p}{1-p}\right)$$

$$\Leftrightarrow e^{z} = \frac{p}{1-p} \Leftrightarrow \frac{1}{e^{z}} = \frac{1-p}{p} = \frac{1}{p} - 1$$

$$\Leftrightarrow \frac{1}{e^{z}} + 1 = \frac{1}{p} \Leftrightarrow e^{-z} + 1 = \frac{1}{p}$$

$$\Leftrightarrow \frac{1}{e^{-z}+1} = p.$$

$$\mathrm{sigmoid}(z) = p = \frac{1}{e^{-z}+1}$$

$$\mathrm{logit}(p) = \ln\left(\frac{p}{1-p}\right).$$

서로 역함수.

softmax Function

- softmax 함수는 sigmoid의 일반형
- Class가 N개인 경우 odd를 이용해 $logit$을 표현하면 다음과 같이 표현할수 있다.
-

$C_1, C_2, \cdots, C_N$ N개의 class에 대하여.

x가 C_k에 속할 확률에 대한 x가 C_i에 속할 승산은 $\exp(z_i) = \dfrac{p(C_i|x)}{p(C_k|x)}$ — ①

x가 C_k에 속할 승산은 $\displaystyle\sum_{j=1}^{k-1} \frac{p(C_j|x)}{p(C_k|x)} = \sum_{j=1}^{k-1} \exp(z_j)$

$$= \frac{1 - p(C_k|x)}{p(C_k|x)} = \sum_{j=1}^{k-1} \exp(z_i)$$

$$\Leftrightarrow \frac{1}{p(C_k|x)} - 1 = \sum_{j=1}^{k-1} \exp(z_i)$$

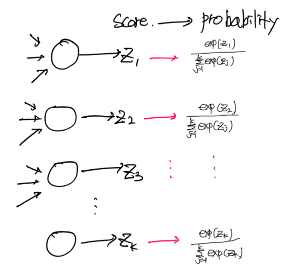
$$\Leftrightarrow p(C_k|x) = \frac{1}{\sum_{j=1}^{k-1} \exp(z_j) + 1} = \frac{1}{\sum_{j=1}^{k-1} \exp(z_j) + \exp(z_k)} = \frac{1}{\sum_{j=1}^{k} \exp(z_j)}$$

$$\therefore p(C_k|x) = \frac{1}{\sum_{j=1}^{k} \exp(z_j)}$$

①에 위의 결과를 대입.

$$\exp(z_i) = p(C_i|x) / p(C_k|x)$$

$$\exp(z_i) / \sum_{j=1}^{k} \exp(z_j) = p(C_i|x)$$



정리

$$P(C_1|x) = y$$

$$P(C_2|x) = 1 - y$$

$$Choose \begin{cases} C_1 & if\ y > 0.5 \\ C_2 & if\ y \le 0.5 \end{cases} \Leftrightarrow \boxed{\frac{y}{1-y}} > 1 \Leftrightarrow \boxed{log_e \frac{y}{1-y}} > 0$$

odds log + odds = **log**istic + probit = **logit**

$$logit(y) = log_e\left(\frac{1}{1-y}\right) = t$$

$$= \frac{y}{1-y} = \exp(t)$$

$$= \frac{1}{y} - 1 = \frac{1}{\exp(t)}$$

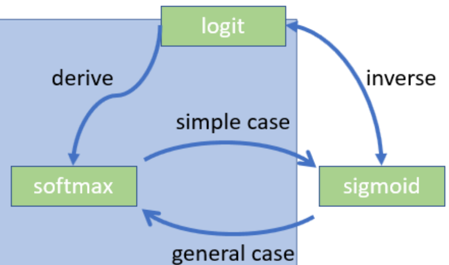
$$= \frac{1}{y} = \frac{1+\exp(t)}{\exp(t)}$$

$$= y = \frac{\exp(t)}{1+\exp(t)}$$

$$= y = \frac{1}{1+\exp(-t)}$$

$$= Sigmoid(t)$$

$$\therefore logit(y) = t,\ \mathrm{Sigmoid(t)} = \mathrm{y}$$



$$C_1, C_2 \cdots, C_K \qquad \frac{P(C_i|x)}{P(C_K|x)} = \exp(t_i)$$

$$\sum_{j=1}^{K-1} \frac{P(C_j|x)}{P(C_K|x)} = \sum_{j=1}^{K-1} \exp(t_j)$$

$$= \frac{1 - P(C_K|x)}{P(C_K|x)} = \sum_{j=1}^{K-1} \exp(t_j)$$

$$= P(C_K|x) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(t_j)}$$

$P(C_i|x) = \exp(t_i) \cdot P(C_K|x)$에 $P(C_K|x)$ 대입

$$= \frac{\exp(t_i)}{1 + \sum_{j=1}^{K-1} \exp(t_i)} = \frac{\exp(t_i)}{\exp(t_K) + \sum_{j=1}^{K-1} \exp(t_j)} \qquad (\because 1 = \frac{P(C_K|x)}{P(C_K|x)} = \exp(t_K))$$

$$\therefore \frac{\exp(t_i)}{\sum_{j=1}^{K} \exp(t_j)} = softmax(t_i)$$