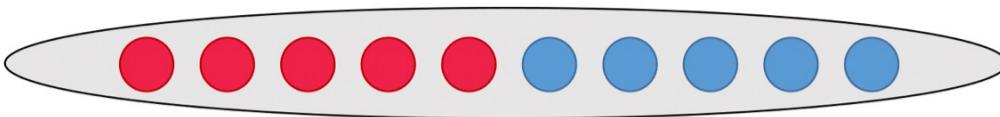


## Question. 8-01

안이 보이지 않는 주머니 속에 빨간 구슬 5개 파란 구슬 5개가 들어있다. 이때, 다음 물음에 답하시오.



- 1) 주머니에서 꺼낸 구슬이 빨간 구슬일 확률과 Odds를 구하시오.
- 2) 빨간 구슬 6개, 파란 구슬 4개가 들어있을 때, 빨간 구슬을 뽑을 확률과 Odds를 구하시오.
- 3) 빨간 구슬 4개, 파란 구슬 6개가 들어있을 때, 빨간 구슬을 뽑을 확률과 Odds를 구하시오.
- 4) 위 과정에서 빨간 구슬을 뽑을 확률이 달라졌을 때 Odds의 값이 어떻게 변화하였는지 설명하시오.
- 5) Odds를 기반으로 Logistic Regression 학습을 진행하는 것이 적절한지 설명하시오.

1) 전체 구슬의 개수가 10개, 빨간 구슬의 개수가 5개이므로,

$$P(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬})} = \frac{5}{10} = 0.5$$

$$O(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬}) - n(\text{빨간구슬})} = \frac{5}{10-5} = 1$$

2) 전체 구슬의 개수가 10개, 빨간 구슬의 개수가 6개이므로,

$$P(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬})} = \frac{6}{10} = 0.6$$

$$O(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬}) - n(\text{빨간구슬})} = \frac{6}{10-6} = 1.5$$

3) 전체 구슬의 개수가 10개, 빨간 구슬의 개수가 4개이므로,

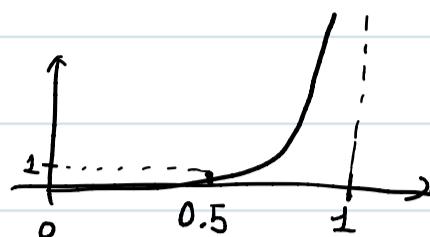
$$P(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬})} = \frac{4}{10} = 0.4$$

$$O(\text{빨간구슬}) = \frac{n(\text{빨간구슬})}{n(\text{전체구슬}) - n(\text{빨간구슬})} = \frac{4}{10-4} = 0.6$$

4)  $\eta$ (빨간구슬) 4 5 6  
 $P(\text{빨간구슬})$  0.4 0.5 0.6  
 $O(\text{빨간구슬})$  0.6 1 1.5  
 $\quad \quad \quad -\frac{1}{3} \quad \quad \quad +\frac{1}{2}$

확률의 변화가 균등하게 나타났음에도 Odds는 불균등하게 값이 변화하였음을 알 수 있다.

5)  $\frac{P(A)}{1-P(A)}$  의 그래프



$$P(A) > P(A^c) : O(A) > 1$$

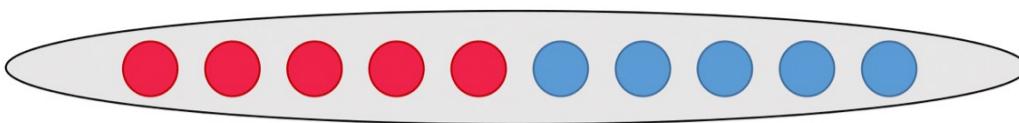
$$P(A) = P(A^c) : O(A) = 1$$

$$P(A) < P(A^c) : 0 < O(A) < 1$$

$P(A) > P(A^c)$ 의 범위와 (1이상의 실수)  $P(A) < P(A^c)$ 의 범위가 (0~1) 달라서 확률의 불균형이 생기므로 Odds를 기반으로 학습을 진행하는 것은 적절하지 않다.

## Question. 8-02

Q8-01과 같이 안이 보이지 않는 주머니 속에 빨간 구슬 5개 파란 구슬 5개가 들어있다. 이때, 다음 물음에 답하시오.



- 1) 주머니에서 꺼낸 구슬이 빨간 구슬일 경우에 대한 Logit을 구하시오.
- 2) 빨간 구슬 6개, 파란 구슬 4개가 들어있을 때, 빨간 구슬에 대한 Logit을 구하시오.
- 3) 빨간 구슬 4개, 파란 구슬 6개가 들어있을 때, 빨간 구슬에 대한 Logit을 구하시오.
- 4) 빨간 구슬의 확률이 달라졌을 때 Logit의 값이 어떻게 변화하였는지 설명하시오.
- 5) Logit을 기반으로 Logistic Regression 학습을 진행하는 것이 적절한지 설명하시오.

$$1) \text{logit}(\text{빨간구슬}) = \ln(O(\text{빨간구슬})) = \ln\left(\frac{P(\text{빨간구슬})}{1-P(\text{빨간구슬})}\right) = \ln\left(\frac{0.5}{0.5}\right) = 0$$

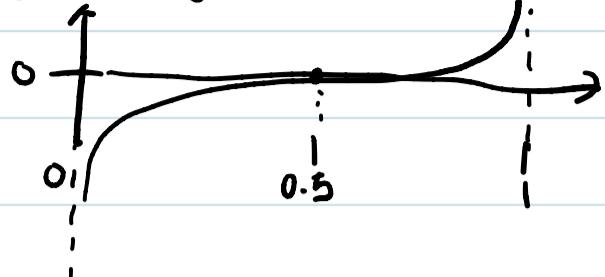
$$2) \text{logit}(\text{빨간구슬}) = \ln(O(\text{빨간구슬})) = \ln\left(\frac{P(\text{빨간구슬})}{1-P(\text{빨간구슬})}\right) = \ln\left(\frac{0.6}{0.4}\right) = \ln\left(\frac{3}{2}\right) \\ = 0.4$$

$$3) \text{logit}(\text{빨간구슬}) = \ln(O(\text{빨간구슬})) = \ln\left(\frac{P(\text{빨간구슬})}{1-P(\text{빨간구슬})}\right) = \ln\left(\frac{0.4}{0.6}\right) = \ln\left(\frac{2}{3}\right) \\ = -0.4$$

4)	$P(\text{빨간구슬})$	0.4	0.5	0.6
	$O(\text{빨간구슬})$	0.6	1	1.5
	$\text{logit}(\text{빨간구슬})$	-0.4	0	0.4

확률의 균등한 변화에 따라 logit 역시 균등하게 변화하였음을 알 수 있다.

- 5)  $\log$ 의 성질에 의해  $\log\left(\frac{P}{1-P}\right) = \log(P) - \log(1-P)$ 로 표현된다.  
 $\log(P)$ 와  $\log(1-P)$ 는 0.5를 기준으로 대칭하므로  
 $\log(P) - \log(1-P)$ 는 0.5를 기준으로 X축에 대칭인 미접인 형태가 된다.



따라서  $P(A) < P(A^c)$ 의 범위와  $P(A) > P(A^c)$ 의 범위의 비중이 같다.  
( 0 이면 실수 ) ( 0 초과 실수 )

$-\infty \sim \infty$ 의 범위를 가지므로 머신러닝에서 Affine function( $\vec{\theta}^T \vec{x}$ )과 같은 역할을 할 수 있으므로 학습에 활용되는 것이 적절하다.

## Question. 8-03

Softmax가 어떻게 Logit을 받아 해당 Class에 대한 Probability를 만들어 낼 수 있는지 보이시오.

1)  $C = \{C_1, C_2, \dots, C_k\}$ , k개의 Class가 있을 때

$\log_{10}(C_i) = \ln\left(\frac{P(C_i)}{P(C_k)}\right)$ 로 표현할 수 있고, 이때  $i \neq k$ 인 i-class의 Logit은  $\ln\left(\frac{C_i}{C_k}\right)$ 이고  $k$ -th class의 Logit은 0이다.

이를 다시  $e^{\log_{10}(C_i)} = \frac{P(C_i)}{P(C_k)}$ 로 표현할 수 있고,  $i \neq k$ ,  $P(C_i) = e^{\log_{10}(C_i)}$ ,  $P(C_k) = e^{\log_{10}(C_k)}$ 이다.  
 $i = k$ ,  $e^{\log_{10}(C_k)} = 1$

이때,  $\sum_{j=1}^{k-1} e^{\log_{10}(C_j)} = \sum_{j=1}^{k-1} \frac{P(C_j)}{P(C_k)} = \frac{1 - P(C_k)}{P(C_k)} = \frac{1}{P(C_k)} - 1$  이 된다.

따라서  $1 + \sum_{j=1}^{k-1} e^{\log_{10}(C_j)} = \frac{1}{P(C_k)} \Rightarrow P(C_k) = \frac{1}{1 + \sum_{j=1}^{k-1} e^{\log_{10}(C_j)}} = \frac{1}{e^{\log_{10}(C_k)} + \sum_{j=1}^{k-1} e^{\log_{10}(C_j)}}$   
 $\Rightarrow P(C_k) = \frac{1}{\sum_{j=1}^k e^{\log_{10}(C_j)}}$  이된다.

$i \neq k$  일 때  $P(C_i) = e^{\log_{10}(C_i)}$ ,  $P(C_k) = e^{\log_{10}(C_k)}$  이므로  $P(C_i) = \frac{e^{\log_{10}(C_i)}}{\sum_{j=1}^k e^{\log_{10}(C_j)}}$  이고

$\sum_{j=1}^k e^{\log_{10}(C_j)}$ 는 고정값,  $e^{\log_{10}(C_i)}$ 는 Logit 입력이 된다.

따라서 Softmax를 통해 Class가 k개인 경우에서 Logit을 입력받아 그 Logit의 대상의 확률을 구할 수 있다.