

기계학습 5주차 과제

12191656 이채연

HomeWorks #1

HomeWorks #1

- Empirical Distribution을 사용하는 이유와 $\mathcal{D} = \{x_1, \dots, x_N\}$ 가 주어졌을 때 $p_{emp}(A)$ 를 어떻게 구하는지 설명 하시오.

Empirical Distribution을 사용하는 이유는 어떤 sample의 고려하고 있는 feature vector의 distribution 정보가 없을 때 아무런 가정을 하지 않고 sample 자체에 대한 feature vector를 가지고 distribution을 empirical하게(경험적, 실험적) 그리기 위해서 empirical distribution을 사용한다.

Dataset $D = \{x_1, \dots, x_N\}$ 가 주어졌을 때 $P_{emp}(A)$ (empirical distribution 또는 empirical measure)를 다음과 같이 구할 수 있다.

$$p_{emp}(A) \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(A)$$

여기서 Dirac measure 인 $\delta_x(A)$ 함수는 다음과 같이 정의된다. x 가 A 라는 우리가 고려하는 데이터 도메인에 있을 때는 1이고, 그러한 데이터 도메인에 있지 않다면 0을 return 하는 delta function 이다.

$$\delta_x(A) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A \end{cases}$$

즉, 어떤 이벤트가 발생이 되면 dirac function은 1이 되고, 아니면 0이 된다. 이렇게 dirac function을 이용해서 표현하면 sample에 대한 histogram을 그릴 수가 있다. 각 sample에 대한 dirac function을 구하고 이를 모두 sum한 뒤, 실제 데이터(sample)의 개수인 N 으로 나눠주면 $P_{emp}(A)$ 를 구할 수 있다. 그러면 모든 data에 동일하게 weight를 할당해서 distribution을 구할 수가 있다. 여기서 weight는 $1/N$ 이 된다.

HomeWorks #2

HomeWorks #2

- $y = f(x) = \begin{cases} 1, & x \text{ is even} \\ 0, & \text{o/w} \end{cases}$. The probability density function $p_x(x)$ of x is uniform on the set $\{1, 2, \dots, 10\}$. Find the $p_y(y)$.

$p_x(x)$ 는 x 에 대한 uniform 한 distribution 이므로 $0.1 = 1/10$ 이다.

y 값이 0 이 나오게 하는 x 가 1, 3, 5, 7, 9 가 있다. $x \in \{1, 3, 5, 7, 9\}$ 일 때 각각의 $p_x(x)$, 즉 $p_x(1), p_x(3), p_x(5), p_x(7), p_x(9)$ 가 모두 0.1 이므로 이를 모두 sum 을 하면 0.5 이다. 따라서 $p_y(0)$ 은 0.5 이다.

y 값이 1 이 나오게 하는 x 가 2, 4, 6, 8, 10 이 있다. $x \in \{2, 4, 6, 8, 10\}$ 일 때 각각의 $p_x(x)$, 즉 $p_x(2), p_x(4), p_x(6), p_x(8), p_x(10)$ 가 모두 0.1 이므로 이를 모두 sum 을 하면 0.5 이다. 따라서 $p_y(1)$ 은 0.5 이다.

정리하면, $p_y(0) = 0.5, p_y(1) = 0.5$ 이다.