

3. Deep Neural Network with MLE

모두 같은 표현

첫번째 식은 seta인 확률분포에서 x sample에 대한 확률

두번째 식은 seta인 확률분포에서 x가 주어졌을 때 y일 확률

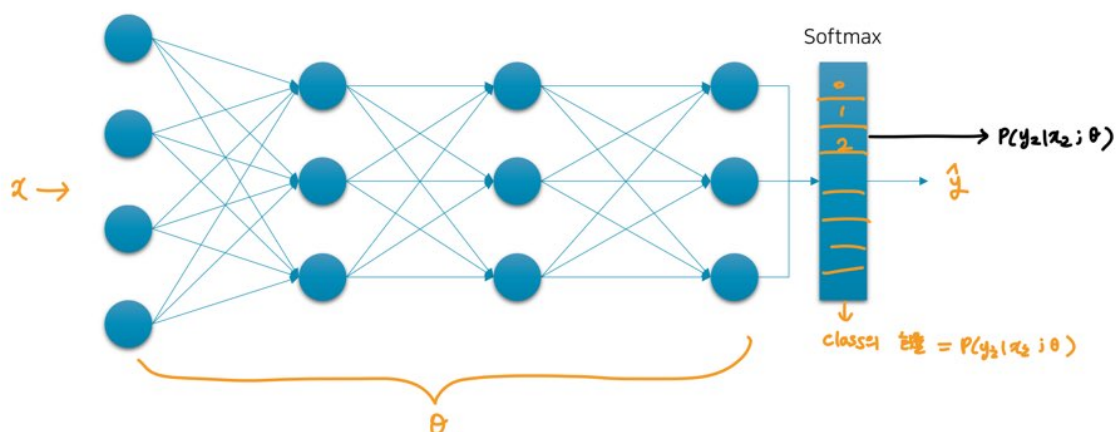
하지만, 표현은 같으나 자세히 들여다보면 값, 확률변수라는 관점이 다르기 때문에 식을 풀 때는 유의해야 하며 여기서는 주로 중간식(값으로 seta를 여김)을 많이 사용

$$P_{\theta}(x) = P(x; \theta) = P(x|\theta)$$

$$P_{\theta}(y|x) = P(y|x; \theta) = P(y|x, \theta)$$

Neural net with MLE

분포 $P(X)$ 로부터 샘플링한 데이터 x 가 주어졌을 때, 파라미터 seta를 갖는 복잡한 비선형 확률분포를 통해서 $P(Y|X, \text{seta})$ 라는 출력에 대한 분포를 얻을 수 있다.



$$P(y|x; \theta), \text{ where } x \sim P(x).$$

이 때, $X=x$ 값이 주어지면 이 분포를 통해서 y 값을 알 수 있기 때문에 분포를 잘 표현하는 파라미터를 찾는 것이 관건! 이 때 파라미터(set)는 각 layer의 weight, bias 값들이다.

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^N -\log P(y_i|x_i; \theta)$$

따라서 Maximum Log Likelihood를 구하는데, 딥러닝은 최소값 찾기에 최적화되어 있으므로 -를 곱해주어 Negative Log Likelihood 값을 최소로 만드는 파라미터를 Gradient Descent를 통해 알아낸다.

$$\theta \leftarrow \theta - \alpha \cdot \frac{\partial \mathcal{L}(\theta)}{\partial \theta}$$