

- 로지스틱 회귀 (Logistic Regression)

로지스틱 회귀는 이진 분류 모델에서 사용되는 분류 모델이다. 이는 대표적으로 Sigmoid 함수를 사용하며 0에 가까운지, 1에 가까운지를 모델링하여 출력을 0 또는 1로 제한한다.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid 함수의 형태는 위와 같으며 로지스틱 회귀의 학습 방법으로는 주로 MLE를 사용한다.

Likelihood 함수는 다음과 같은 형태인데, $L(\theta|x) = \prod p_i^{y_i} (1 - p_i)^{1-y_i}$ 이 양변에 log를 취하여 활용하면 다음과 같다. $\log L(\theta|x) = \sum (y_i \log p_i + (1 - y_i)(1 - p_i))$ 이는 negative log가 아닌 그냥 log이므로 log를 취한 likelihood의 최댓값을 구해도 likelihood 함수의 최댓값을 구하는 것과 동일하다. 따라서 최댓값을 만들어주는 θ 를 추정하도록 하여 학습한다. 이외에도 경사 하강법을 사용하는 방식으로 학습을 하는 방법도 존재한다.

- 결정 트리 (Decision Tree)

Decision tree는 특정 기준에 따라 데이터를 구분하는 모델로, 한 번의 분기 때마다 영역을 두 개로 구분한다. 이를 과도하게 많은 질문으로 나눠 오버피팅이 되는 것을 피하기 위해서 노드당 최소 데이터 수를 정해주는 기법도 존재한다. 결정 트리의 주요 개념으로는 불순도가 존재한다. 불순도는 해당 척도 내에 서로 다른 데이터가 얼마나 섞여있는지를 나타낸 것으로 이를 나타내는 대표적인 척도가 엔트로피이다. 엔트로피는 $Entropy = - \sum_i (p_i) \log_2(p_i)$ 이와 같은 공식으로 높을수록 불순도가 높다는 뜻이다. 분기 이전의 엔트로피에서 분기 이후의 엔트로피를 뺀 것을 정보 획득이라 하고 이를 최대화하는 방향으로 학습을 진행한다.

- 서포트 벡터 머신(Support Vector Machine)

SVM은 N차원을 N-1 차원으로 나눌 수 있는 초평면을 찾는 기법이다. 이는 margin을 기준으로 해서 초평면을 구하는데 margin은 클래스 말단에 위치한 데이터들 사이의 거리를 의미한다. 기법의 이름에 있는 서포트 벡터는 이들 중 margin에 가장 가까이 위치한 데이터이다. $\min_w \frac{1}{2} ||w||^2$ 이러한 공식으로 찾는데, 이처럼 오차없이 나눌 수는 없으므로 어느정도 오분류를 허용하는 $\min_w \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i$ 이러한 공식을 사용한다. 이 경우는 선형으로 분리가 가능한 경우이고, 이외의 비선형의 경우에는 공간을 확장하는데, 이 때 계산량을 줄이고자 커널트릭이라는 기법을 사용하여 확장된 것처럼 사용하여 계산량을 줄인다.

- K-최근접 이웃(K-Nearest Neighbors)

K-최근접 이웃 기법은 입력 데이터에 대해서 학습 데이터 중 가장 가까운 K개를 찾고 그 이웃들의 레이블을 기반으로 예측하는 기법이다. 우선 $d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2}$ 이와 같은 유클리드 거리와 같은 방식으로 입력 데이터와 학습 데이터 사이의 거리를 계산한다. 이 후에 정한 거리에 따라 가장 가까운 K개를 선택한다. 그 후 분류의 경우에는 $\hat{y} = \operatorname{argmax}_c \sum_{i \in N_K(\mathbf{x})} \mathbf{1}(y_i = c)$ 이처럼 가장 많이 등장한 label을 결과로 반환하고 회귀의 경우에는 $\hat{y} = \frac{1}{K} \sum_{i \in N_K(\mathbf{x})} y_i$ 이와 같이 평균값을 예측 결과로 사용한다.

- 나이브 베이즈(Naïve Bayes)

나이브 베이즈는 베이즈 정리를 기반으로 하는 각 특성이 독립이라고 가정하고 계산을 진행하는 모델이다. 기반으로 하는 베이즈 정리의 공식은 다음과 같다. $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ 이를 바탕으로 각 label별로 조건에 맞는 조건부 확률을 계산하여 가장 큰 확률을 가지는 label을 예측값으로 고르는 단순한 형태이다. 이는 간단하고 빠르지만 조건이 독립이 아닌 경우에 대해서 부정확할 수 있고, 연속형 데이터나 불균형한 데이터 등에 처리가 어려운 단점이 존재한다.

- 인공신경망(Artificial Neural Network)

인공신경망은 뇌의 뉴런 구조를 모방한 모델로 여러 층과 노드를 통해서 예측값을 구하는 기법이다. 층은 입력층, 은닉층, 출력층 3가지로 구분할 수 있고, 각각 데이터가 세팅되는 층, 드러나지 않고 가려져 있는 층, 예측값이 나타나는 층이다. 노드들은 매우 많은 가중치와 bias의 조합으로 이루어져 있고 결과 값에서 오차가 발생하면 다시 앞쪽으로 가중치를 갱신하는 Backpropagation이 존재한다. 입력값은 transfer function을 거치고 activation function을 거쳐서 다음 노드로의 출력값으로 계산된다. Activation function으로는 앞에서 나온 sigmoid 함수도 쓰이고,

$\max(x, 0)$ 이러한 형태를 가진 ReLU 함수도 쓰인다. 마지막 출력층 전에서 출력 함수를 통해서 출력 값이 계산되는데 대표적으로는

$$g\left(w^{(L)\top} h^{(L)} + b^{(L)}\right) = \left[\frac{e^{w_1^{(L)\top} h^{(L)} + b_1^{(L)}}}{\sum_{k=1}^K e^{w_k^{(L)\top} h^{(L)} + b_k^{(L)}}}, \dots, \frac{e^{w_K^{(L)\top} h^{(L)} + b_K^{(L)}}}{\sum_{k=1}^K e^{w_k^{(L)\top} h^{(L)} + b_k^{(L)}}} \right]^\top$$

이러한 Softmax 함수가 존재한다. 이

들과 손실함수를 사용하여 계산과 역전파를 반복하여 누적하여 출력값을 계산해내는 것이 인공신경망이라고 할 수 있다.