**로지스틱회귀분석(Logistic Regression)**

**데이터 분할과 성과분석**

**Theory**

데이터마이닝에서 분류분석으로 모형을 개발하면 임의의 모형(Random Model)보다 더 우수한 분류 성과를 보이는지와 고려된 서로 다른 모형 중 어느 것이 가장 우수한 예측 및 분류 성과를 보유하고 있는지 확인해야 한다. 모형의 성능을 평가하기 위해서 데이터를 분할하여 모형의 분류 및 예측 정확도를 **오분류표(Confusion Matrix), ROC(Receiver Operation Characteristic) 그래프, 이익도표** 등으로 성과분석을 수행한다.

**로지스틱회귀분석(Logistic Regression)**

**Theory**

로지스틱 회귀모형은 반응변수가 범주형인 경우에 적용되는 회귀모형이다. 이 방법은 새로운 설명변수의 값이 주어질 때 반응변수의 각 범주에 속할 확률이 얼마인지를 추정하여, 추정확률을 기준치에 따라 분류하는 목적으로 활용된다.

로지스틱 회귀가 선형 회귀와 다른 점은 학습을 통해 선형 함수의 최적선을 찾는 것이 아니라 시그모이드(Sigmoid) 함수 최적선을 찾고, 해당 함수의 반환값으로 확률을 구해 분류를 결정한다는 것이다.

로지스틱 회귀는 데이터의 정규 분포도에 따라 예측 성능의 영향이 미치므로, 정규 분포 형태의 표준화를 거치는 것이 좋습니다. 또한, 주요 파라미터로는 “penalty”와 “C”가 있습니다. **사실상 규제를 하지 않을 경우 별도로 하이퍼파라미터 튜닝을 할 필요도 없다.**

**오즈비(Odds ratio)**

로지스틱 회귀모형은 오즈의 관점에서 해석이 가능하다. 오즈는 성공할 확률이 실패할 확률의 몇 배인지를 나타내는 확률이며, 오즈비는 오즈의 비율이다.

**▷ 표준화 -> 로지스틱 회귀**