

지도학습

- X를 사용해 Y를 예측할 때, 학습데이터에 X, Y 데이터 모두 존재
 - X는 독립변수(피쳐), Y는 종속변수(라벨)
 - 비지도학습은 X만 존재.
- 크게 회귀 모델과 분류 모델로 나뉨.
 - 회귀 : 라벨이 연속형
 - 분류 : 라벨이 범주형

회귀분석

- 독립변수와 종속변수(연속형) 사이의 관계를 추정하는 통계 기법
 - 독립변수 값 변화에 의한 종속변수 값 변화 예측.
- 일반적으로 잔차

선형 회귀 분석은 4가지 가정이 충족시켜야함

- 선형관계
- 잔차 독립성
- 정규성
- 등분산성

OLS

- OLS는 선형 회귀에서 사용되는 방법 중 하나입니다.
- 목표는 잔차(예측과 실제 값 간의 차이)의 제곱을 최소화하여 모델의 가중치를 결정하는 것입니다.
- 최소자승법을 사용하여 회귀 계수를 추정하며, 이 방법은 잔차의 합을 최소화하는 회귀 계수를 찾습니다.

라쏘 회귀

- 라쏘 회귀는 선형 회귀의 한 변종으로, L1 규제를 사용합니다.
- 목표는 잔차의 합과 L1 규제 항의 합을 최소화하는 회귀 계수를 찾는 것입니다.
- L1 규제는 일부 계수를 정확히 0으로 만들어 변수 선택(Feature Selection)의 역할을 수행할 수 있습니다.

파라미터	파라미터 설명	입력값
alpha	L1 규제 페널티 계수	{float, default=1.0}
fit_intercept	절편 포함 여부 결정	{bool, default=True}
max_iter	이테레이션 횟수의 상한선	{int, default=1000}
tol	조기 종료에 대한 허용 오차	{float, default=1e-4}
positive	파라미터 제약 조건	{bool, default=False}
selection	계수의 업데이트 방식	{cyclic, random, default='cyclic'}

- 라쏘 모델 : LassoCV와 LassoLarsCV 클래스로 교차검증으로 라쏘의 알파 하이퍼파라미터를 선택하는 모델 선택법 수행

릿지 회귀

- 릿지 회귀도 선형 회귀의 한 형태로, L2 규제를 사용합니다.
- 목표는 잔차의 합과 L2 규제 항의 합을 최소화하는 회귀 계수를 찾는 것입니다.
- L2 규제는 모든 계수를 0에 가깝게 만들어 과적합을 방지하는 데 도움을 줍니다.

파라미터	파라미터 설명	입력값
alpha	L1 규제 페널티 계수	{float, default=1.0}
fit_intercept	절편 포함 여부 결정	{bool, default=True}
max_iter	이테레이션 횟수의 상한선	{int, default=1000}
tol	조기 종료에 대한 허용 오차	{float, default=1e-4}
solver	최적화 기법 선택	{'auto', 'svd', 'cholesky', 'lsqr', 'sparse_cg', 'sag', 'saga', 'lbfgs'}
positive	파라미터 제약 조건	{bool, default=False}

- 릿지 모델 : RidgeCV 클래스로 교차검증으로 라쏘의 알파 하이퍼파라미터를 선택하는 모델 선택법 수행

로지스틱 회귀

- 로지스틱 회귀는 분류 문제에 사용되는 선형 모델입니다.
- 목표는 입력 변수의 선형 결합을 로지스틱 함수에 적용하여 이진 또는 다중 클래스 분류를 수행하는 것입니다.
- 로지스틱 함수는 S자 형태로, 0과 1 사이의 값을 출력합니다.

파라미터	파라미터 설명	입력값
penalty	규제에 사용된 기준을 지정	{l1, l2, elasticnet, none}
multi_class	다중클래스분류 문제의 상황에서 어떤 접근 방식을 취할지 결정	{문자열, 'ovr', 'multinomial', default='auto'}
C	규제 강도	{float, 'default'= 1.0}
solver	최적화문제를 풀기 위한 알고리즘	{'newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga, default='lbfgs'}
max_iter	solver에 의해 진행되는 수렴을 위한 반복의 최대 횟수 지정	{int, default=100}
class_weight	학습 시 클래스에 따라 가중치를 다르게 주고 싶을 때 사용할 수 있는 파라미터	{dict, 'balanced', default=None}

- 로지스틱 모델: LogisticRegressionCV 클래스를 통해 규제 강도의 역수 C 등 회귀의 파라미터를 탐색