**ㄱ**

**- 교차분석/카이제곱 검정(범주형 자료 분석)**

: 카이제곱 값은 편차의 제곱값을 기대빈도로 나눈 값들의 합니다. 적합도 검정(분포가 고른지), 독립성 검정(가설검정), 동질성 검정(편향된 데이터인 경우) 3가지로 분류 가능하다.

**ㄹ**

**- 랜덤 포레스트(앙상블)**

: 의사결정나무는 분산이 크므로 배깅과 부스팅보다 더 많은 무작위성을 주어 약분류기를 생성 후 이를 선형결합하여 최종 학습기를 만드는 방법이다. 부트스트랩 데이터를 추출하고, 이들을 훈련 후 대표 변수 샘픙을 도출하여 리프노드로 분류한다. 그리고 이 노드들의 선형 결합으로 최종 모델을 결정한다. 트리의 개수, 최대 깊이, 임의성 이 주요 파라미터이다.

**ㅂ**

**- 배깅(앙상블)**

: 학습 데이터에서 다수의 부트스트랩 자료를 생성하고 각각의 자료에 대해 모델을 생성하여 최종 다수결로 분류하는 기법

**- 범주형 자료 분석**

: 분할표 분석, 교차분석(카이제곱 검정), 피셔의 정확검정 등이 있다.

**- 부스팅(앙상블)**

: 잘못 분류된 개체들에 가중치를 적용, 약분류기를 결합해 강분류기를 만든다. 동일 가중치 분류기로 시작해서 가중치 변경을 이어간다.

**- 부트스트랩(배깅)**

: 동일한 크기의 표본을 랜덤복원추출로 뽑은 자료. 100개의 샘플을 추출하더라도 한번도 선택되지 않는 데이터가 36.8%가 된다.

**- 분할표 분석(범주형 자료 분석)**

: 일반적으로 행은 독립변수, 열을 종속변수로 배치한다. 특히 2원 분할표를 많이 사용. 상대위험도, 오즈비로 활용한다.

**ㅇ**

**- 아리마 모델/자기회귀누적이동평균모형**

: 비정상 시계열 모형으로, AR이나 MA, ARMA 등으로 정상화할 수 있다. p(AR모형), q(MA모형), d(ARMA로 정상화할 때 몇 번 차분했는지)라는 차수를 가지고 있다.

**- 앙상블**

: 여러가치 모형들의 예측/분류 결과를 종합하여 최종적 의사결정에 활용하는 기법. sampling(배깅, 부스팅)과 variable(랜덤 포레스트)로 나눌 수 있다.

**ㅈ**

**-주성분 분석**

: 상관관계가 있는 고차원 자료를 최대한 보존하면서 저차원 자료로 변환하는 차원축소 방법. 기존의 상관성이 높은 변수들을 요약, 축소한다. 고윳값이 높은 순서대로 고유벡터만으로 데이터를 복원한다.