

01

ADP 시험 분석

[수업 목표]

ADP취득을 위해서는 ADP 시험이 어떻게 나오는지 철저하게 분석해야 합니다. 이번 시간에는 ADP에 대해 분석하고 철저히 준비하는 방법에 대해 알아보아요!

21회 기출 문제 분석

1. 머신 러닝 (50점)

학생 성적에 관한 데이터셋 제공 약 350행짜리 소규모 데이터.

1-1. 시각화 포함 탐색적 자료분석(EDA)

1-2. 결측치 식별하고 결측치를 예측하는 두 가지 방법 정도를 쓰고, 선택한 이유를 설명.

1-3. 범주형 변수 인코딩이 필요한 경우를 식별하고, 변환을 적용하시오. 선택한 이유를 설명.

1.4. 데이터 분할 방법을 2가지 쓰고 적절한 데이터 분할을 적용. 선택한 이유 설명.

1.5. svm, xgboost, randomforest 3개의 알고리즘 공통점을 쓰고 이 예측 분석에 적합한 알고리즘인지 설명.

1.6. 세 가지 모델 모두 모델링 해보고 가장 적합한 알고리즘 선택하고 이유 설명. 한계점 설명하고 보완 가능한 부분 설명. 현장에서 사용시 주의할 점 등에 대해 기술.

[21회 기출 문제 분석_머신러닝]

0:32

이유를 설명하는 시험이므로 철저한 준비가 필요하겠죠?

21회 기출 문제 분석

2. 통계분석 (50점)

1. 연속형 독립변수 여러개의 소규모 데이터. (총 29점)

변수명은 순서대로 $x_1 \sim x_{10}$, 의미는 없는 데이터

1-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 $rmse$ 구하시오

1-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오.

α 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 $rmse$ 를 계산

1-3. 라쏘 회귀로 2-2과 동일한 문제

2. 독립변수 하나 종속변수 하나 소규모 데이터 다항회귀(12점)

다항 회귀를 3차까지 적용하고 각 차수별 데이터포인트 스캐터 플롯과 계수와 기울기 선을 그리세요.

3. ANOVA분석 (9점)

변수 3개(하나는 범주형 변수/ 나머지 두 개는 수치형 연속변수)

이원분산분석을 수행하고 통계표를 작성하시오.

[21회 기출 문제 분석_통계분석]

7:34

크게 머신러닝 파트와 통계 분석 파트가 나뉘죠?

빠르게 먹으면 체합니다. 머신러닝과 통계 분석을 하기 전에

우리는 데이터를 핸들링하고 전처리 하는 방법을 숙달해야 합니다

22회 기출 문제 분석

1. 기계학습(50점)

1.데이터 탐색(당뇨 데이터 세트, 데이터에 헤더가 없음, 대신 시험지에 변수명 제공)

1.1.1 탐색적 데이터 분석 수행하시오(시각화 포함)

1.1.2 이상치 처리하시오

1.1.3 앞선 두 단계에서 얻은 향후 분석시 고려사항 작성

2. 클래스 불균형을 처리하시오

2.1 업 샘플링 과정 설명하고 결과 작성

2.2 언더 샘플링 과정 설명하고 결과 작성

2.2 둘 중 선택하고 이유 설명

3. 모델링 하시오

3.1 최소 3개 이상 알고리즘 제시하고 정확도 측면의 모델 1개와 속도 측면의 모델 1개를 꼭 구현(총 2개 이상)

3.2 모델 비교하고 결과 설명

3.3 속도 개선을 위한 차원 축소 설명하고 수행, 예측 성능과 속도 비교하고 결과 작성

[22회 기출 문제 분석_머신러닝]

10:20

22회 기출 문제 분석

2. 통계분석 (50점)

1. 금속 성분 함유량 변수 1개. (1월 데이터) 제품에 금속 재질 함유량의 분산이 1.3을 넘으면 불량이라고 보고 제조사별로 차이가 난다고 제보를 받는다. 분산에 대해 검정을 수행하시오.

1.1 연구가설과 귀무가설 작성 / 1.2 양측 검정 / 1.3 검정통계량, 가설 채택

2. Lot별 불량 제품 수량 데이터. lot 번호와 불량제품수 두 개의 열. 각 lot별 200개에 대한 불량제품 수.

2.1 p관리도에 따라 관리중심선(center line), 관리 상한선, 하한선 구하시오

2.2 관리도 시각화 하시오

3. 표에 제품 1,2를 만드는데 사용되는 재료 a b c 컬럼 있고 재료에 따라 최종 만들어지는 제품 두 개에 대한 수량 있다.

제품 수량을 최대로 뽑으면서 수익이 최적이 되도록 작성하시오.(10점)

4. 상품 a와 b가 있을 때 다음과 같은 구매 패턴이 있다고 함. aa bb bbbb aa aaa bb bbb aa bb a b

4.1 구매하는 패턴으로 봐서 두 상품이 연관이 있는지 가설 세우고 검정하시오

4.2 연구가설 귀무가설 세우시오

4.3 가설 채택하시오

[22회 기출 문제 분석_통계분석]

14:19

ADP 시험 유형1.

[통계분석 + 시각화]

<p>[15회, 2019-12-14] 데이터 처리 및 통계 분석 - timestamp 처리 / date 기준 데이터 병합 - h:mm, A/B/C/D/E, 전역 사용량 데이터 - yyyymmdd, 평균 기준</p> <p>1. 아래 형태의 데이터 생성 : 3개 데이터를 date기준으로 병합 필요 → yyyy-mm / A/B/C/D/E / 사용량 2. 요일 변수 생성 및 A/B/C/D/E별 평균 사용량 출력, 그래프 출력 3. 요일간 사용량 분석을 수행하고 가장 차이가 있는 요일 도출</p>	<p>[17회, 2020-06-21] 통계 분석 (설명데이터 분석) → (사전에 역문항들에 대한 처리 필요) 1. 그룹별 평균, 표준편차, 왜도, 첨도 산출</p> <p>시각화 및 시계열 분석 (코로나 데이터) (20점) 1. 전체 인구대비 누적 사망률이 가장 높은 5개 국가 추출 후, 국가별 일일확진자, 누적확진자, 일일사망자, 누적사망자 시계열 그래프 출력 2. 위험지수 생성 및 해석 3. 시계열 분석 및 예측 모델 생성</p>	<p>[21회, 2021-06-09] 2. 연속형 독립변수 여러개의 소규모 데이터. 변수명은 순서대로 x1~x10 이라 의미 없음 2-1. 데이터 8:2로 분할하고 선형회귀 적용하시오. 결정계수와 mse 구하시오 2-2. 데이터 8:2로 분할하고 릿지 회귀 적용하시오. alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 mse를 계산 2-3. 라쏘 회귀로 2-2과 동일한 문제 3. 독립변수 하나 종속변수 하나 소규모 데이터. 다항 회귀를 3차까지 적용하고 각 차수별 데이터포인트 스케터 플롯과 계수와 기울기 선을 그리라고 함(12점). 4. ANOVA 분석(9점) 변수 3개 하나는 abcd 각각을 값으로 갖는 범주형 변수 나머지 두 개는 수치형 연속변수 이원분산분석을 수행하고 통계표를 작성하시오.</p>
2019	2020	2021

[ADP 시험유형 분석]

18:29

[다음 수업 예고]

다음 시간에는 ADP보다는 쉽지만 ADP를 준비하면서 취득할 수 있는 빅데이터 분석기사를 분석해보겠습니다.