

데이터과학자로 가는길

서울시립대학교 통계데이터사이언스

2022. 12



교과과정

어떤 과목을 어떤 순서대로 들을까?



Carriers

- 데이터룰 다루는 현업에서의 업무

- 데이터를 수집, 저장, 가공하는 일 (데이터 엔지니어링)

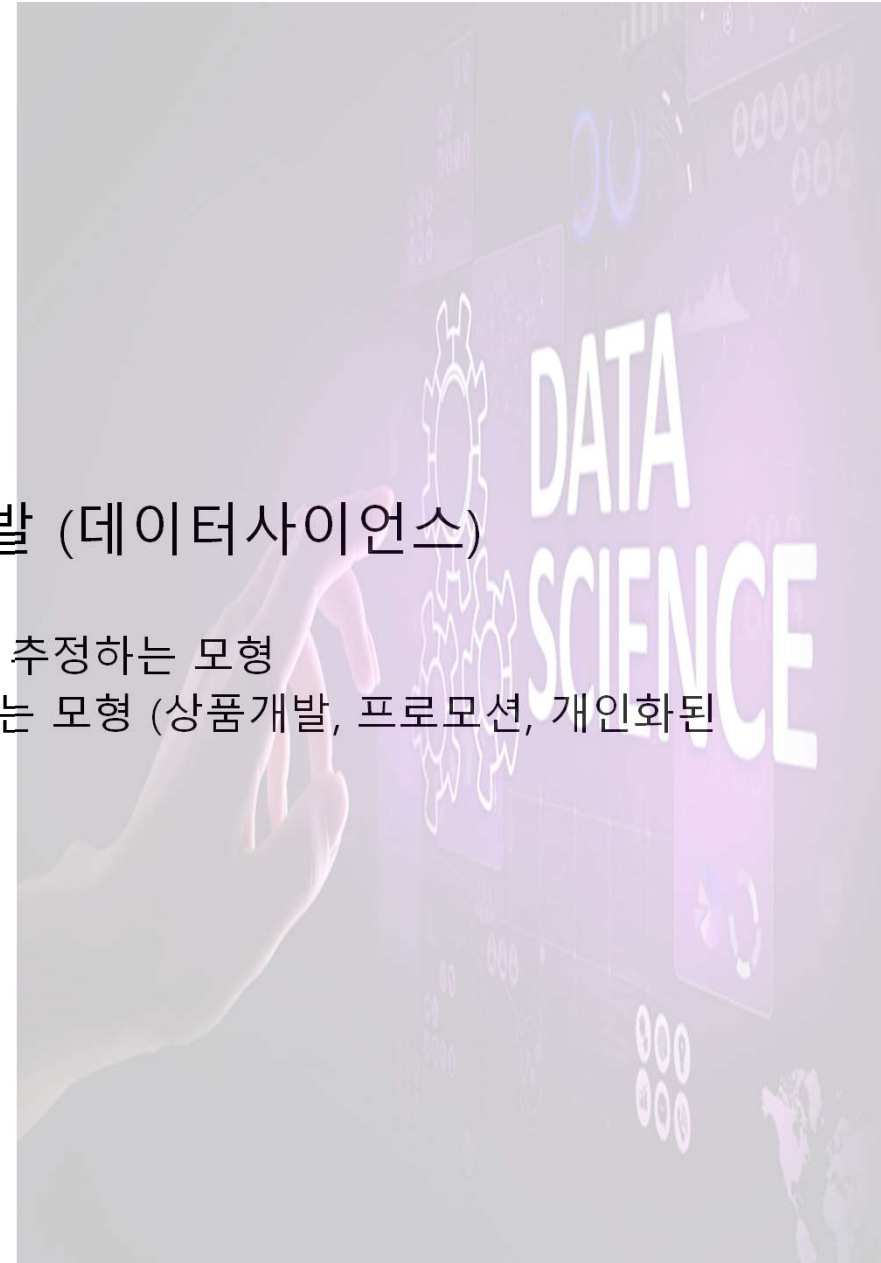
- 예)

- 어떤 데이터가 stream 형태로 전달되고 있음 (2억개/일)
 - 이 데이터를 분석하기 위해서는 컴퓨터에 지속적으로 저장하고 관리해야함 (DB, 컴퓨터 통신관련 지식이 필요: SQL (MySQL, Oracle, Postgre, MongoDB, Marian DB 등)
 - 데이터에 대해 잘 알고 있는가? (데이터의 적용분야에 대한 배경지식)
 - 데이터를 모두 저장할 것인지, 필요한 부분만 저장할 것인지?
 - 정형화되지 않은 데이터를 어떻게 잘 저장할 것인지?
 - 센서와 컴퓨터간 통신을 어떻게 관리할 것인지?
 - 컴퓨터의 부하를 조절하기 위해 필요한 의사결정들...



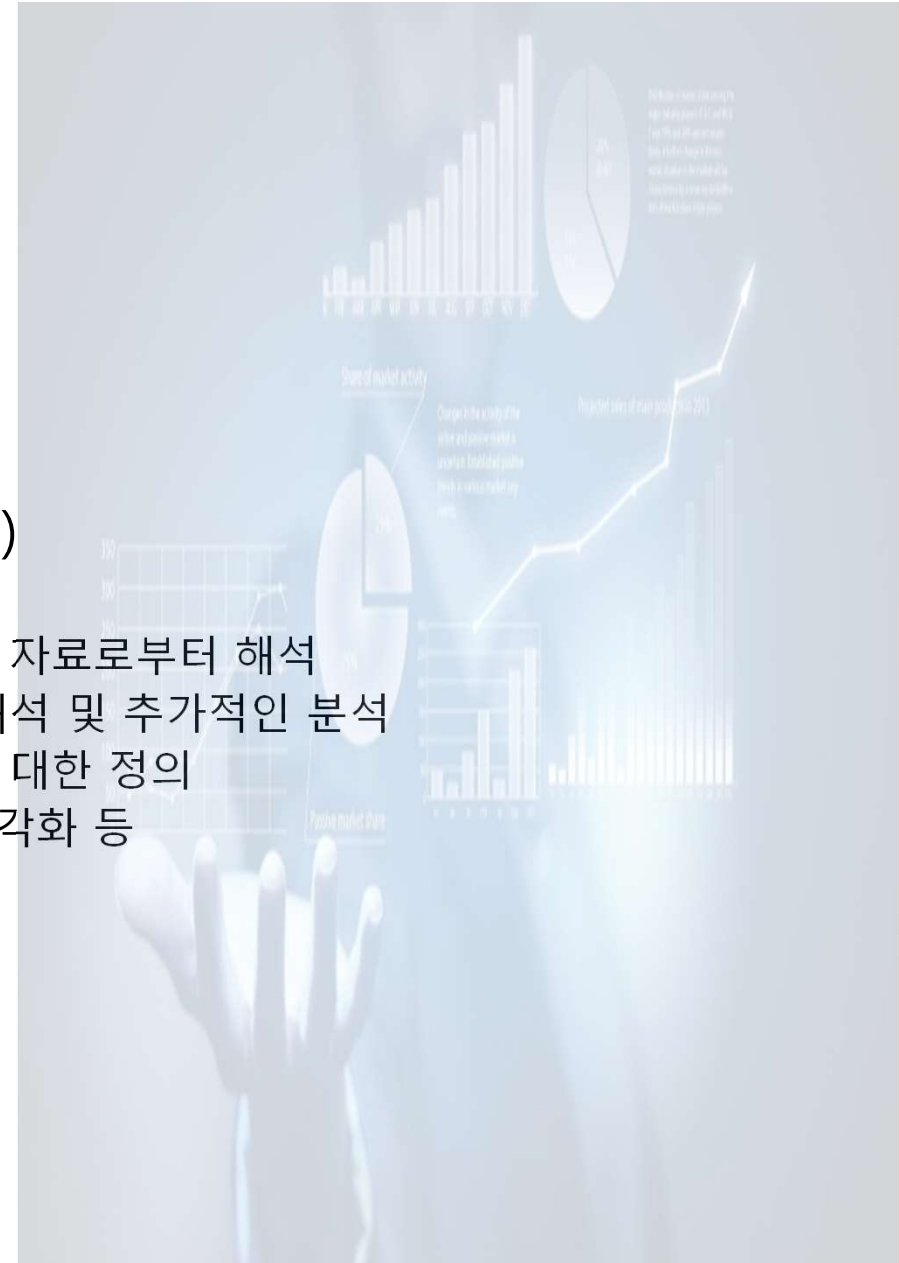
Carriers

- 데이터를 다루는 현업에서의 업무
 - 데이터를 이용한 예측 모형, 설명 모형 개발 (데이터사이언스)
 - 예)
 - 고객의 구매 데이터를 웹사이트 재방문 확률을 추정하는 모형
 - 유사한 상품을 구매하는 고객의 군집을 추정하는 모형 (상품개발, 프로모션, 개인화된 광고)
 - 항적의 이동 경로의 예측
 - ... 너무 많음



Carriers

- 데이터를 다루는 현업에서의 업무
 - 데이터의 분석 및 해석 (데이터애널리틱스)
 - 예)
 - 고객의 구매 데이터를 웹사이트 재방문 패턴을 자료로부터 해석
 - 유사한 상품을 구매하는 고객의 군집에 대한 해석 및 추가적인 분석
 - 작업장 사고 데이터를 통해 사고의 위험요인에 대한 정의
 - 연차별 성과목표 달성도와 기업역량에 대한 시각화 등



Undergraduate Curriculum

- General Courses (Undergraduate)
 - (수리)통계학/확률
 - 통계모형
 - 프로그래밍/계산



데이터과학자 (Core)

1-1 1-2 2-1 2-2 3-1 3-2 4-1 4-2

전필

전선

통계학개론
및 실습

기초확률론

통계자료처리
및 실습

수리통계학

응용확률론

행렬과 대수
I

행렬과 대수
II

통계수학

프로그래밍
언어 및 실
습

회귀분석 I

통계데이터
베이스

회귀분석 및
실습 II

시계열분석
및 실습

다변량통계
학 및 실습

정보이론

통계계산

기계학습 및
실습

딥러닝

수리통계학
II

데이터시각
화

이산자료 분
석

실험계획

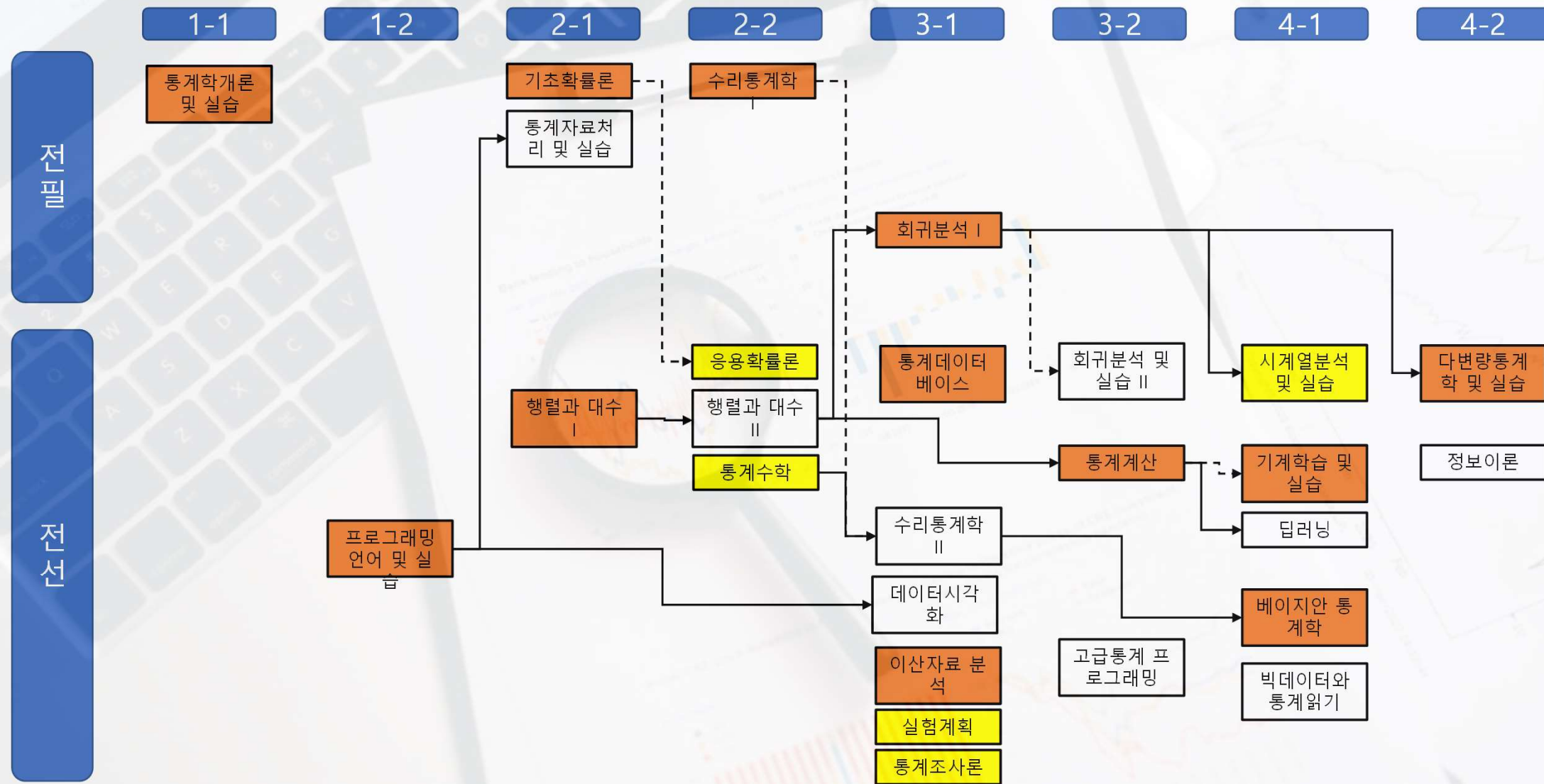
통계조사론

고급통계 프
로그래밍

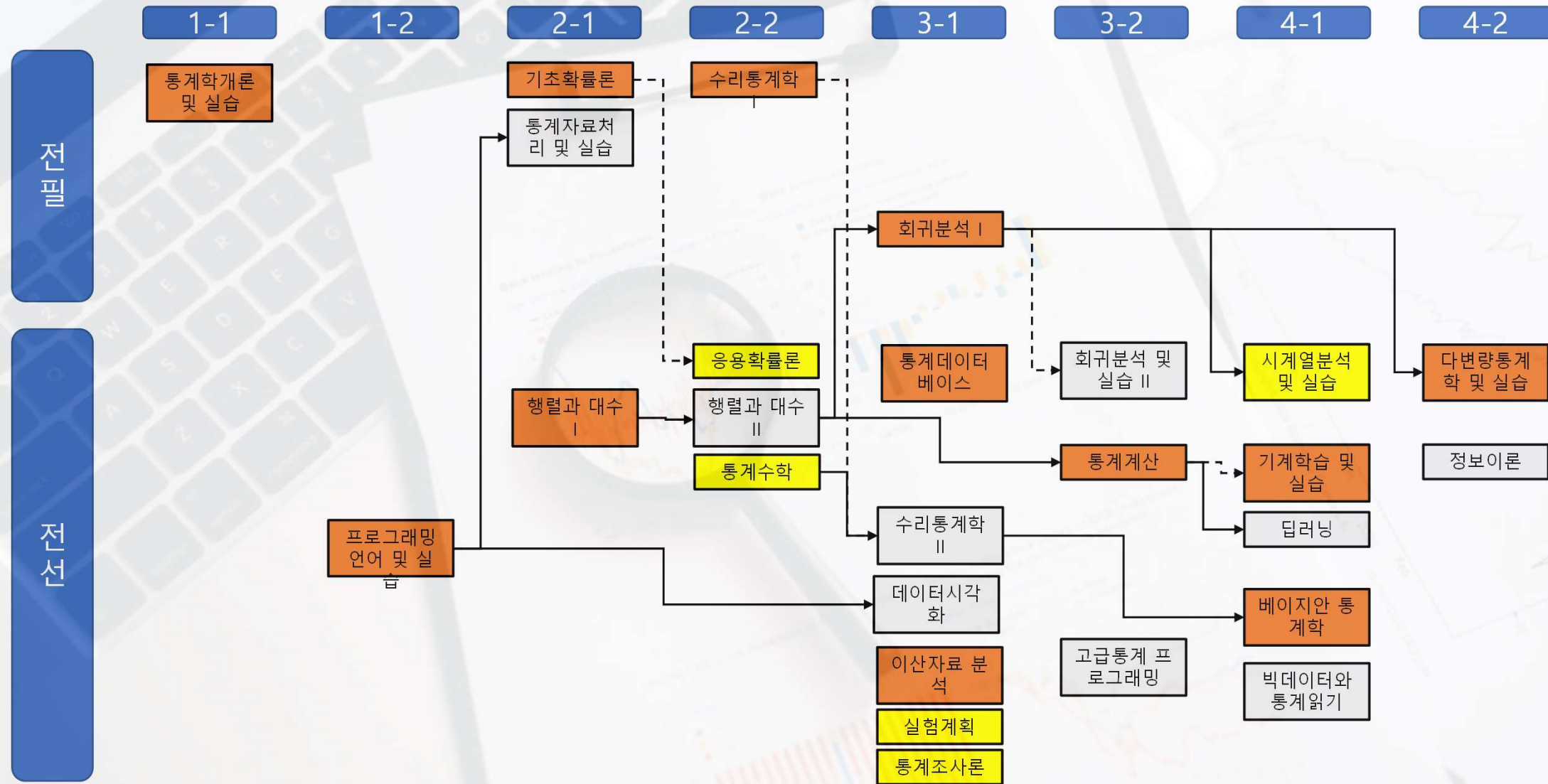
베이지안 통
계학

빅데이터와
통계읽기

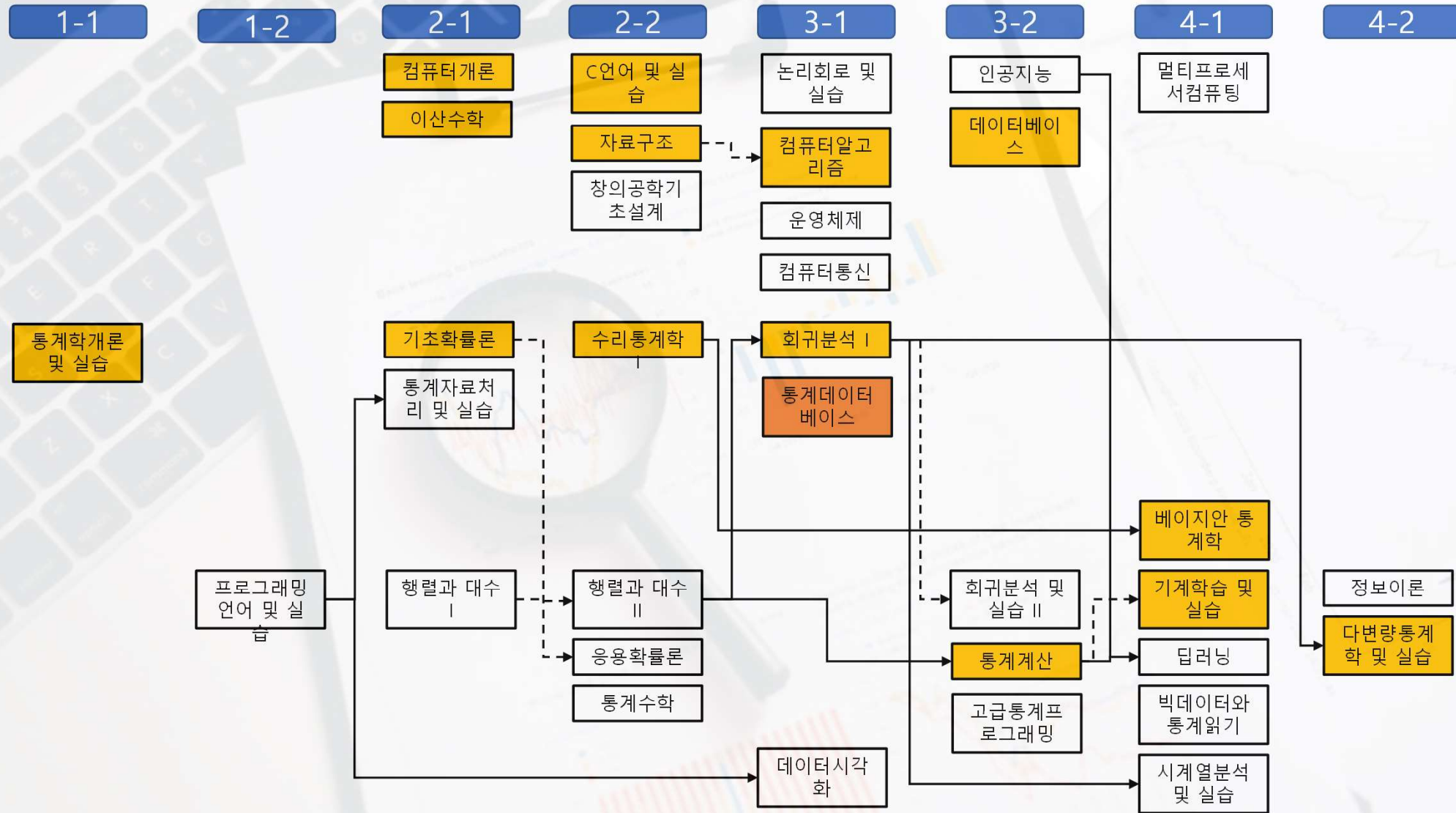
데이터과학자 (Core+)



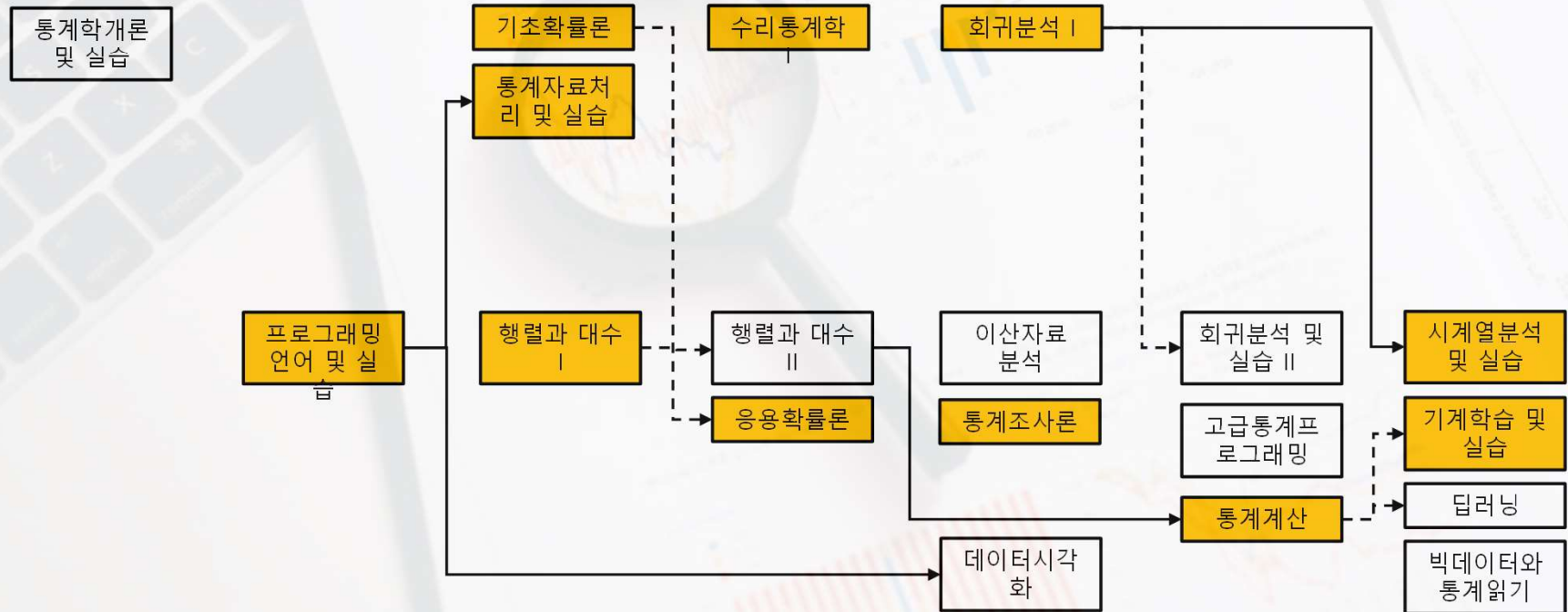
데이터과학자 (Expert)



데이터엔지니어 (Core)



데이터분석가(예시: 금융분야)



데이터분석가(예시: 보험계리사)

1-1 1-2 2-1 2-2 3-1 3-2 4-1 4-2

복수전공

전필

전선

경제학원론 I(미시)

미시경제학

경제학원론 II(거시)

거시경제학

경제수학

회계원리

국제금융론

국제무역론

계량경제학

계리모형론

재무관리

특화분야 배경지식 (Domain)

통계학개론
및 실습

기초확률론

수리통계학

회귀분석 및
실습 I

프로그래밍
언어 및 실
습

행렬과 대수
I

응용확률론

행렬과 대수
II

수리통계학
II

보험통계학

베이지안 통
계학

보험통계학
II

의학통계학

베이지안 통
계학

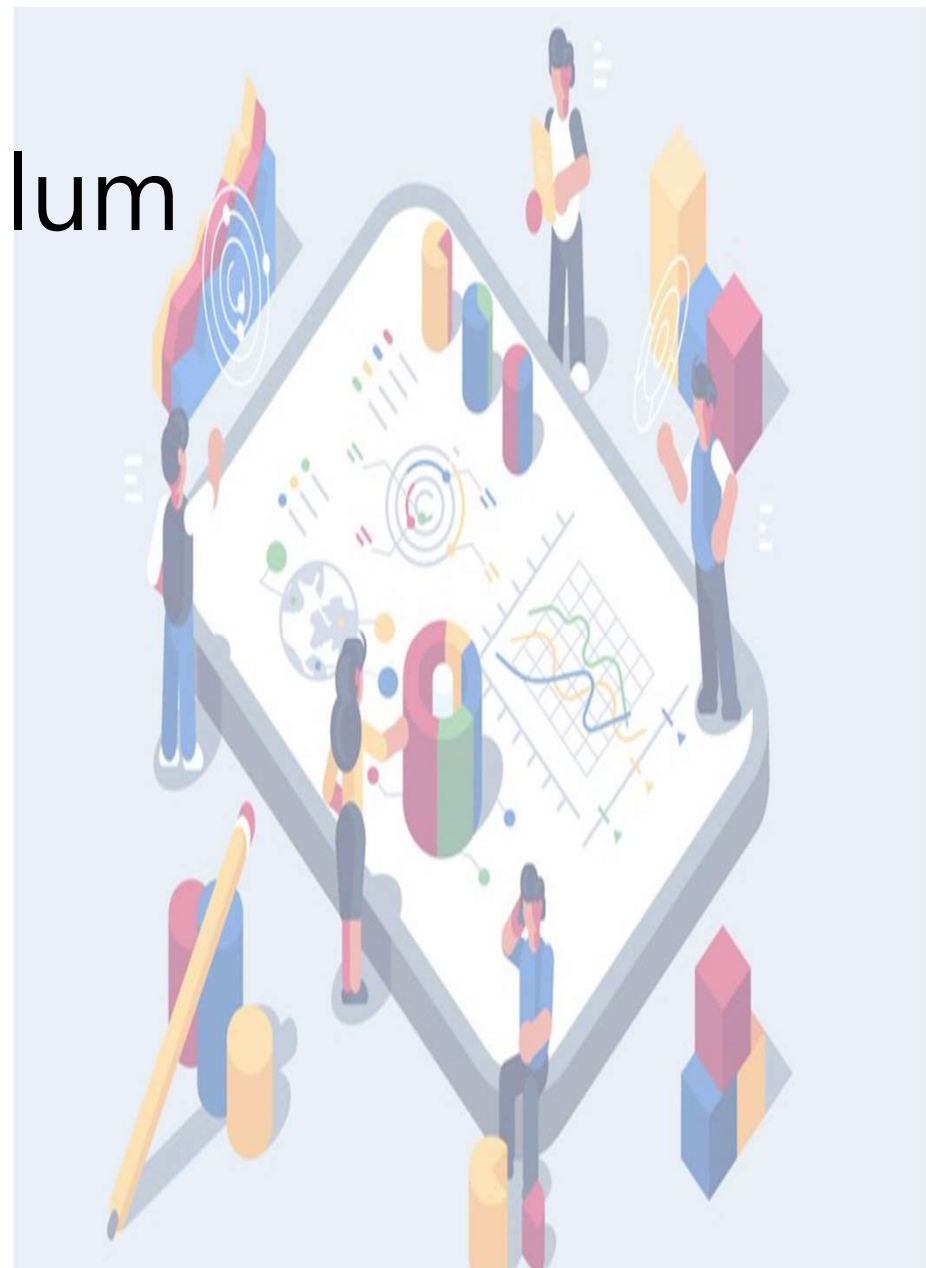
시계열분석
및 실습

금융통계

통계계산

Graduate Course Curriculum

- 공통역량 (대학원 과정)
 - (수리)통계학/확률
 - 수리통계학
 - 확률론, 확률과정론
 - 표본론
 - 고차원통계학
 - 베이시추론.....
 - 통계모형
 - 프로그래밍/계산
 - 전문분야의 선택
- (수리)통계학/확률
 - 수리통계학
 - 확률론, 확률과정론
 - 표본론
 - 고차원통계학
 - 베이시추론.....
- 통계모형
 - 선형모형
 - 응용통계학
 - 기계학습
 - 딥러닝
 - 구조형자료분석.....
- 프로그래밍/계산
 - 통계계산
 - 최적화
 - 확률최적화





Education/Research Support

대학원을 가면 뭐가 좋나?

RESEARCH AREA



김규성

표본론

연구분야 : 표본론, 통계조사론
연구실 : 미래관 701호
전화번호 : 02-6490-2631
Homepage :
E-mail kskim@uos.ac.kr 주소복사



박창이

기계학습

연구분야 : 통계적 학습, 비모수적 추론
연구실 : 미래관 704호
전화번호 : 02-6490-2634
Homepage : <http://statlearn.uos.ac.kr>
E-mail park463@uos.ac.kr 주소복사



전종준

기계학습

연구분야 : Probability model for ranking
연구실 : 미래관 712호
전화번호 : 02-6490-2637
Homepage : <http://ranking.uos.ac.kr>
E-mail jj.jeon@uos.ac.kr 주소복사



안수한

응용확률

연구분야 : 확률과정론, 큐잉이론
연구실 : 미래관 715호
전화번호 : 02-6490-2632
Homepage :
E-mail sahn@uos.ac.kr 주소복사



이용희

의학통계

연구분야 : Variance component and mixed
연구실 : 미래관 714호
전화번호 : 02-6490-2635
Homepage : <http://sampling.uos.ac.kr/~yl>
E-mail ylee@uos.ac.kr 주소복사



최연진

기계학습

연구분야 :
연구실 : 미래관 711호
전화번호 : 02-6490-2639
Homepage : <https://sites.google.com/site/yunjincho..>
E-mail ycstat@uos.ac.kr 주소복사



정병철

패널데이터

연구분야 : 선형모형
연구실 : 미래관 703호
전화번호 : 02-6490-2633
Homepage :
E-mail bcjung@uos.ac.kr 주소복사



김성곤

응용확률

연구분야 : 응용확률, 큐잉이론
연구실 : 미래관 713호
전화번호 : 02-6490-2636
Homepage :
E-mail sgkim@uos.ac.kr 주소복사



박재휘

데이터베이스

연구분야 : 데이터베이스, 정보검색, 자연어 처리
연구실 : 미래관 702호
전화번호 : 02-6490-2630
Homepage :
E-mail jaehui@uos.ac.kr 주소복사

RESEARCH AREA (2023) with Students



김규성

표본론

연구분야 : 표본론, 통계조사론

연구실 : 미래관 701호

전화번호 : 02-6490-2631

Homepage :

E-mail kskim@uos.ac.kr 주소복사

- 1) 성향점수를 이용한 통계무응답 처리
- 2) 매칭을 통한 자료 통합방법 연구
- 3) 준가중치(pseudo-weight)를 이용한 비확률표본 활용방법론



안수한

응용확률

연구분야 : 확률과정론, 큐잉이론

연구실 : 미래관 715호

전화번호 : 02-6490-2632

Homepage :

E-mail sahn@uos.ac.kr 주소복사

- 1) 확률론
- 2) 큐잉이론



정병철

패널데이터

연구분야 : 선형모형

연구실 : 미래관 703호

전화번호 : 02-6490-2633

Homepage :

E-mail bcjung@uos.ac.kr 주소복사

- 1) 패널회귀모형 을 이용한 패널데이터(panel data) 분석
- 2) 다양한 형태의 포아송 및 음이항 회귀모형을 이용한 계수자료에 대한 분석
- 3) 일반화선형모형을 이용한 범주형 자료에 대한 분석

RESEARCH AREA (2023) with Students



박창이

기계학습

연구분야 : 통계적 학습, 비모수적 추론

연구실 : 미래관 704호

전화번호 : 02-6490-2634

Homepage : <http://statlearn.uos.ac.kr>

E-mail park463@uos.ac.kr 주소복사

- 1) 커널(kernel)에 의한 분류방법
- 2) 정규화(regularization) 기반의 변수선택(feature selection)
- 3) 계수 데이터(count data)에 대한 그래프 모형(graphical model)



이용희

의학통계

연구분야 : Variance component and mixed models

연구실 : 미래관 714호

전화번호 : 02-6490-2635

Homepage : <http://sampling.uos.ac.kr/~ylee/>

E-mail ylee@uos.ac.kr 주소복사

- 1) 계층모형(반복측정자료, 반복측정 등)에 대한 계산법 고찰 및 프로그램 구현
- 2) 개인정보보호를 위한 인공지능 모형을 이용한 재현자료(synthetic data) 생성 실습
- 3) 의학연구에서의 자료 분석 방법 고찰 및 실습



김성곤

응용확률

연구분야 : 응용확률, 큐잉이론

연구실 : 미래관 713호

전화번호 : 02-6490-2636

Homepage :

E-mail sgkim@uos.ac.kr 주소복사

- 1) 이분산 시계열 자료 분석(예, 주가, 환율 등의 시계열)을 통한 위험값의 계산.
- 2) 공간 시공간 자료 분석

RESEARCH AREA (2023) with Students



전종준

기계학습

연구분야 : Probability model for ranking

연구실 : 미래관 712호

전화번호 : 02-6490-2637

Homepage : <http://ranking.uos.ac.kr>

E-mail jj.jeon@uos.ac.kr 주소복사

- 1) Transformer를 이용한 장기 시계열 모형 예측
- 2) 가상자산 (crypto asset) 최적 포트폴리오 모형 개발
- 3) 인과추론 리뷰



최연진

기계학습

연구분야 :

연구실 : 미래관 711호

전화번호 : 02-6490-2639

Homepage : <https://sites.google.com/site/yunjincho..>

E-mail ycstat@uos.ac.kr 주소복사

- 1) interpretable AI
- 2) Community detection
- 3) Selective Inference



박재휘

데이터베이스

연구분야 : 데이터베이스, 정보검색, 자연어 처리

연구실 : 미래관 702호

전화번호 : 02-6490-2630

Homepage :

E-mail jaehui@uos.ac.kr 주소복사

- 1) 언어 모델을 활용한 Visual Question Answering 응용
- 2) 언어 모델을 활용한 Recommendation System 응용
- 3) GAN 기반 Style Transfer 응용

Opportunity

• 연구

- 석사연구실 (공간, 컴퓨터)
- 계산자원 (통계학과 서버운용)
- 정기세미나 참석
- 학회지 논문 (포스터) 발표지원 (국내, 해외)
- 학생인건비 (지원가능)
 - 박사과정의 안정적이고 장기적인 연구지원을 위한 노력



연구실 환경 (2022.11.10)

• 프로젝트

- 교수 연구실별로 다양한 프로젝트의 수행
 - 서울시, KCB, 신한카드, SKT, 결합데이터 분석
 - 해양 다중 신호처리와 위치추정
 - 가상자산 평가 및 포트폴리오 위험 추정 등



2022.8.1-3 2022 한국인공지능학회 (제주도, 신화월드 컨벤션)



2022.7 인과추론 세미나 (10시간 과정)

Opportunity

- 기대효과

- 연구분야를 정함으로써 전문분야에 대한 심화연구가 가능함.
- 수업/학회참여활동/프로젝트 참여를 통해 데이터관련 현업의 업무를 직접 접하고 문제를 해결할 수 있는 기회
 - 이력서
 - 코딩테스트
 - 실무면접
- 안정적인 재정지원
 - 특별한 경제활동을 하지 않고 학업을 수행할 수 있는 재정 지원이 가능함

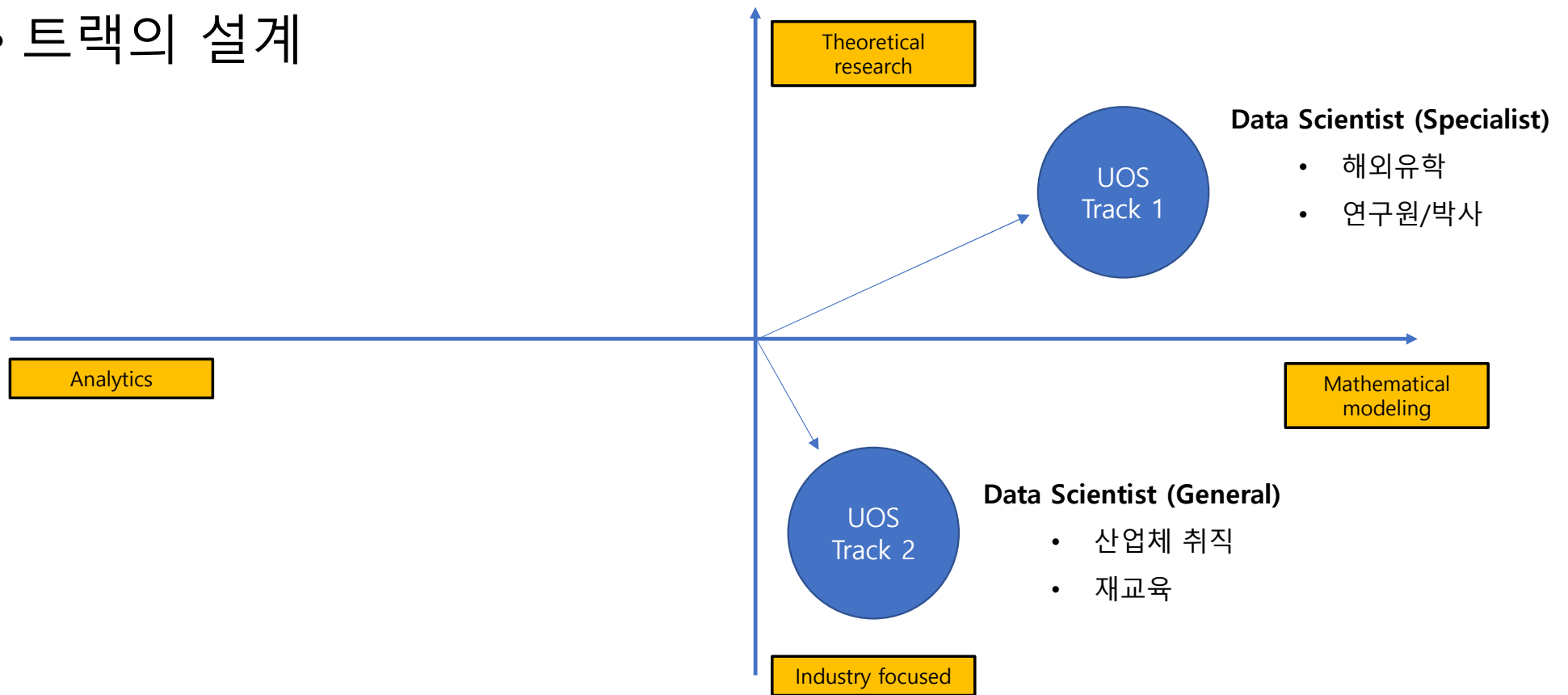


도시빅데이터AI연구원 계산자원 환경

디지털혁신인재양성 사업 (빅데이터)- 10.8억/년 (5년)
데이터사이언스융합인재양성사업 - 10억/년 (6년)
기초연구실 - 5억/년 (향후 2.5년)
교수별 개인기초연구사업 및 프로젝트

Positioning

• 트랙의 설계



Admission

- 학생미래설계학기

- 방학중 학생미래설계학기(연구학기) 신청

- 계절학기로 3학점 인정 (연구실 생활을 미리 접할 수 있는 기회!) 최대 6학점까지 전공인정 가능
 - 신청 전 희망 지도교수와 상담 -> 지도교수 강좌개설 신청 -> 선정시 방학중 개설 가능 (교수 1인당 최대 학생 2명)
 - 학생: 연구활동비 30만원 지원

- 학석사 연계과정

- 복수전공생의 경우 해당 학과 학과장 승인 필요
 - 학석 연계과정 지원시 전공분야를 확정할 필요는 없음

- 학석박 연계과정

- 석사 진학후 박사 과정 연계 (학석박의 경우 최대 수강학기를 1.5년 단축 가능)
 - 전공분야 및 지도교수를 정하는 것을 추천





서울시립대학교
UNIVERSITY OF SEOUL

서울시립대학교
통계데이터사이언스 대학원에서
데이터과학자의 꿈을 이루세요!
- U Our Star -