

데이터사이언스학과 (2017-2학기 신설)

※ 개별 홈페이지는 제작 중

I 개황

- 데이터 과학 및 인공지능 관련 사회.산업적 요구 증가
 - 최근 4차 산업혁명이라고 불릴 만큼 데이터를 기반으로 한 인공지능이 주요 기술/산업 분야로 대두되고 있으며 이에 따라 전 세계적으로 관련 역량을 보유한 인력에 대한 요구가 확대되고 있음
 - CNN은 2012년 최고 유망 신규 직종으로 데이터과학자 선정
 - Google Trend에서는 전통적인 통계학자의 수요를 데이터과학자가 대체할 것으로 전망
- 우리 대학 내의 관련 교육과정을 통합을 통한 내실화
 - 수학과, 산업공학과, 전자공학과, 소프트웨어학과, 금융공학과, 의학과, 미디어학과 등의 학과에 다수의 관련 전문가를 확보
 - 교내 관련 분야 교원들이 대학원 학과를 중심으로 유기적인 협력관계를 만들고 공동으로 교육과정을 운영하는 과정을 통하여 교육을 내실화

[Contacts]

- 학과장 : 최영우 교수(☎ youngwoo@ajou.ac.kr, ☎ 031-219-2570, Facebook ychoi100)
- 기타 문의
 - 입시, 등록, 장학 : ☎ 031-219-2302
 - 수업, 성적, 졸업, 논문 : ☎ 031-219-2303

II 교육목적

“4차 산업혁명 시대를 선도할 산업체 맞춤형 데이터사이언스 전문 인재 양성”

- 데이터사이언스 분야 전문 역량을 갖춘 인재
 - 프로그래밍 기술, 분석 알고리즘, 비즈니스 컨설팅 기술 등에 대한 전문 역량 확보
 - 데이터 분석 기술들을 창의적으로 응용하여 효율적이고 유의미한 결과를 유도
- 데이터사이언스 분야 글로벌 경쟁력을 갖춘 인재
 - 데이터 분석 기술 분야의 기술 발전을 주도
 - 도메인 특화 데이터 분석 전문성을 통한 국내 산업의 글로벌 경쟁력 강화에 기여

III 학생 진로 및 수요 전망

□ 데이터사이언스 관련 해외 인력 시장 동향

- 21세기 ‘가장 매력적인’ 직업으로 데이터 과학자 선정 (Harvard Business Review, 2012, 10)
- 미국에서는 2018년까지 140,000-190,000명의 데이터 분석 전문가와 1,500,000명의 데이터 분석 기반 관리자가 필요(맥킨지 보고서, 2011. 05)
- 데이터 분석 기반 의사결정 시스템을 갖춘 기업이 2014년 약 12%에서 2015년 27%로 증가하였고 향후 42%이상의 기업이 시스템을 구축할 예정 (2015 글로벌 시장조사기관 (IDG)에 의한 전세계 1,139개 기업 대상 설문조사)

□ 국내 업계의 인력 수요 전망

- 국내 유수의 대기업(삼성전자, SK telecom 등)에서는 데이터 사이언스 관련 전문 팀을 운영 중이며 다수의 조직에서 신규 인력 채용에 있어 인력 난을 호소하고 있음
- 수년 전부터 국내에도 데이터 사이언스 관련 전문 기업들(천명소프트, 휴민텍, 인퍼니트 헬스케어, 다음소프트, 오픈메이트, 빅스터 등)이 속속 등장하였고, 활발히 활동 중이며 고정적으로 인력을 채용하고 있음

IV 학과장

직 급	성 명	학위(학위수여대학)	전공분야
교수	최영우	박사(University of Wisconsin)	해석학

V 교수진

직 급	성 명	학위(학위수여대학)	전공분야	비 고
교수	신현정	박사(서울대)	Data Mining, Machine Learning	
부교수	구형일	박사(서울대)	전기컴퓨터공학	
부교수	손경아	박사(미·Carnegie Mellon University)	데이터마이닝(Data mining)	
조교수	한경식	박사(Pennsylvania State University)	Human Computer Interaction, Social Computing	
부교수	김강석	박사(Indiana University Bloomington)	컴퓨터공학	
교수	신현준	박사(한국과학기술원)	컴퓨터그래픽스	
조교수	권순선	박사(서울대학교)	통계학	
조교수	김하영	박사(Purdue Univ-West Lafayette)	Stochastic Analysis and Mathematical Finance	
교수	박래웅	박사(충북대 의대)	의료정보학	
조교수	윤덕용	박사(아주대 의대)	의학	
임용예정	정재훈	박사(Brown Univ)	Applied Math	

VI 교육과정

□ 교육과정표

학수구분	과목코드	과목명	학점	시간	비 고
전공선택	DS500	데이터사이언스개론	3	3	
	DS520	데이터처리언어	3	3	
	IIS6611	고등데이터마이닝	3	3	산업공학과 공동개설
	IIS644	고등통계분석	3	3	산업공학과 공동개설
	ECE641	패턴인식론	3	3	전자공학과 공동개설
	IIS680	기계학습	3	3	산업공학과 공동개설
	CSE6111	고급인공지능	3	3	컴퓨터공학과 공동개설
	DS525	데이터컴퓨팅	3	3	
	DS531	소셜미디어분석	3	3	
	FEN612	계산금융	3	3	금융공학과 공동개설
	MES6035	의생명 데이터 패턴분석과 데이터마이닝 I	3	3	의생명과학과 공동개설
	DS501	빅데이터사례연구 (종합설계)	3	3	
	DS502	데이터사이언스특강	1	2	
	MATH653	이론통계	3	3	수학과 공동개설
	MATH654	응용통계	3	3	수학과 공동개설
	ECE741	컴퓨터비전	3	3	전자공학과 공동개설
	ECE631	확률및스토캐스틱프로세스	3	3	전자공학과 공동개설
	MED993	의학통계학	3	3	의생명과학과 공동개설
	CSE6110	고급데이터마이닝	3	3	컴퓨터공학과 공동개설
	DS751	의료정보분석특론	3	3	
	MATH681	수학적 모델링I	3	3	수학과 공동개설
	DS722	기계학습특론	3	3	
	MATH671	산업수학특론I	3	3	수학과 공동개설
	MATH672	산업수학특론II	3	3	수학과 공동개설
	DS560	현장실습1	3	6	
	DS561	현장실습2	3	6	
	DS562	현장실습3	3	6	

□ 이수학점

구분	석사	박사	통합
전공학점	24	30	48
연구학점	6	15	15
총 학점	30	45	63

□ 수료 및 졸업요건  자세한 내용은 대학원 홈페이지 참조

[수료요건]

- 최소학기 등록(석사:4학기, 박사:4학기, 통합:8학기)
 - * 학석사연계과정 1학기 단축가능, 통합과정 2학기 단축가능
- 교육과정 이수 완료(이수학점 충족)
- 누계평점평균 3.0 이상

[졸업요건]

- 학위과정 수료(예정)자가 학위청구논문 제출 후 심사통과
 - * 학위청구논문 제출자격 요건
 - 비교과교육 이수 : 연구윤리, 연구노트, 논문작성법
 - 자격시험 통과 : 외국어시험, 종합시험
 - 학회지게재 요건 충족(박사 및 통합과정)

VII 교과목 개요

• DS500 데이터사이언스개론 (Introduction to Data Science)

빅데이터 시대의 요구에 맞게 부상하고 있는 데이터 사이언티스트의 역할과 필요 역량에 대한 이해, 데이터 사이언스의 상세분야 등 데이터 사이언스의 개요를 다룬다. 또한, 데이터 사이언스 전반에서 다루는 다양한 분야를 소개하는 것을 목적으로 한다. 따라서, 다양한 전공분야에서 데이터 사이언스의 지식을 가진 교수들의 윤강으로 진행된다. 학습 내용은 다음과 같다.

- (i) 데이터 사이언스 프로세스 관련 핵심 개념 - 데이터의 이해, 데이터 분석 도구, 데이터 활용
- (ii) 데이터 사이언스를 적용한 사례 등을 소개 - 다양한 분야에서 활용되고 있는 데이터 사이언스 적용사례, 텍스트, 이미지, 그래프 등 상이한 종류의 데이터들을 소개

• DS520 데이터처리언어 (Computer Programming for Data Processing)

컴퓨터를 활용하여 데이터를 처리하기 위한 프로그래밍 기술을 학습한다. 특히 프로그램 개발하기 위한 변수와 형, 조건, 반복, 함수 등에 대한 기본 개념과 더불어 이러한 개념을 Python, C++, JavaScript 등 다양한 컴퓨터 프로그래밍 언어에 적용하기 위한 방법을 학습한다. 학습 결과로 학생들은 프로그래밍에 관한 기본 개념과 간단한 프로그래밍 개발 방법, 프로그램을 작성하여 데이터를 처리하기 위한 방법을 이해할 수 있다.

• IIS6611 고등데이터마이닝 (Advanced Data Mining)

Data capture has become inexpensive and ubiquitous as a by-product of innovations such as the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, intelligent machines, and the amount has been increasing at an incredible rate due to technological advances. "Data mining" refers to a collection of techniques for extracting "interesting" relationships and knowledge hidden in a mountain of data in order to assist managers or analysts to make intelligent use of them. A number of successful applications have been reported

in areas such as credit rating, fraud detection, database marketing, customer relationship management, and stock market investments. In this course, we will examine a variety of data mining techniques evolved from the disciplines of statistics and artificial intelligence (or machine learning), and practice them in recognizing patterns and making predictions from an applications perspective. Application (or case) surveys and hands-on experimentations with easy-to-use software will be provided.

• **IIS644 고등통계분석 (Advanced Statistics)**

공학에 필요한 수리통계학을 강의하는 과목이다. 주요 논제로는 확률론, 확률변수와 분포, 변수변환, 샘플링분포, 추정, 검정론 등이다. 본 교과에서는 학부에서 수강한 통계학을 이론적으로 접근하며, 통계적 접근방법론의 이해를 배양함을 목표로 한다.

• **ECE641 패턴인식론 (Pattern Recognition Theory)**

패턴인식 방법들에 대해 공부한다. 먼저 비지도학습과 지도학습 등의 개념과 이들의 차이점에 대해서 공부하고, 지도학습 중에서도 분류 문제와 리그레션 문제가 어떻게 다른지에 대해서 공부한다. 각 방법들의 대표적인 알고리즘들과 이들의 수학적 모델링에 대해서 다룬다. 학기말에는 얼굴인식 시스템의 구현 등 기말 프로젝트를 수행하면서 패턴인식의 노하우를 깨치게 된다.

• **IIS680 기계학습 (Machine Learning)**

Machine learning is all about finding generalized patterns from data. The whole idea is to replace the "human writing code" with a "human supplying data" and then let the system figure out what it is that the person wants to do by looking at the examples. In recent years, many successful applications of machine learning have been developed, ranging from data-mining

programs that learn to detect fraudulent credit card transactions, to autonomous vehicles that learn to drive on public highways. At the same time, there have been important advances in the theory and algorithms that form the foundation of this field. The goal of this class is to provide an overview of the state-of-art algorithms used in machine learning and different perspectives, and hopefully to gain some understanding of what's going on the next. We will discuss both the theoretical properties of these algorithms and their practical applications.

• **CSE6111 고급인공지능 (Advanced Artificial Intelligence)**

지식표현과 추론을 집중적으로 공부한다. 특히 Ontology Engineering을 위한 지식 표현 및 추론을 중심으로 공부하며, 이에 대한 사례연구 중심으로 심도 있게 다룬다.

• **DS525 데이터컴퓨팅 (Data Computing)**

대용량 데이터를 처리하고 기계학습을 수행하기 위한 병렬 컴퓨팅의 개념과 응용 능력을 학습한다. 특히 GPU를 이용한 병렬 계산의 개념과 활용 방법, 복수의 컴퓨터를 활용하기 위한 하둡 등의 설치 및 응용 프로그램 개발 방법을 학습하고 실제 응용에 적용하기 위한 프로젝트를 수행한다.

• **DS531 소셜미디어분석 (Social Media Analysis)**

1. API 활용 및 웹 크롤링을 통한 소셜미디어 데이터 수집 및 저장
2. 데이터의 전처리, 압축 및 correlation, regression, and classification을 위한 다양한 기법을 활용하여 분석
3. 언어학적 특징 분석 및 감성 분석
4. 연구 목적을 위한 다양한 툴을 활용한 소셜미디어 데이터 분석 및 시각화

• **FEN612 계산금융 (Computational Finance)**

기존에 학습했던 파생상품의 가격결정, 가치측정, 위험 측정 및 관리, 헷징 시뮬레이션 등을 C++, JAVA, Matlab, Excel VBA 등 프로그래밍 언어를 이용하여 직접 수행해본다. 이를 위해 프로그래밍의 기초에 대해 학습을 하고 다양한 방법으로 파생상품 평가와 리스크 측정하는 원리를 배운다. 이를 위해 분석적 방법과 수치해석 방법에 대해 학습하고, Value at Risk와 Greek에 대해 학습한다.

• **MES6035 의생명 데이터 패턴분석과 데이터마이닝 (Pattern Analysis and Data Mining about Biomedical Data I)**

• **DS501 빅데이터사례연구(종합설계) (Big Data Case Study)**

IT, 의료, 금융 등에 관련된 빅데이터 분석 사례들을 공부하고, 각자 관심 도메인의 실제 빅데이터 분석이 필요한 문제들을 찾고, 찾은 문제에 대해서 데이터 획득, 데이터 정제, 모델링, 결과 분석까지 해보는 빅데이터 종합설계 과정임.

• **DS502 데이터사이언스특강 (Thermodynamic of solids)**

현재 데이터사이언스 분야에서 관심 주제들에 대해 자세히 다룹니다. 현재 관심 주제들에 대해 자세히 공부하고, 자신의 관심 주제에 적용해 볼 수 있습니다. 현재 관심 주제들을 공부함으로써 실제 산업에 필요한 지식을 학습하게 됩니다.

• **MATH653 이론통계 (Theory of Statistics)**

확률 및 통계와 수리통계학에서 취급되지 않았던 통계이론들 중 통계적 추정이론과 검정이론에 해서 다룬다. 분포족, 충분성에 대하여 다루고, 최소오차 추정량과 최대우도추정량과 이를 계산하는 방법에 대하여 다룬다. 최강력검정이론과 불편검정이론을 학습하고, 최대가능도법에 의한 검정이론을 다룬다.

• **MATH654 응용통계 (Applied Statistics)**

자료를 모형화하는 방법으로 선형모형에 대해 다룬

다. 모형선택, 추정, 모형검정과 같은 방법론에 대해서 단순회귀, 다중회귀, 분산분석 등의 모형에 기반하여 최소제곱방법을 이용한 추론을 다룬다. 추가적으로 우도에 관한 추론도 다루고자 한다. 선형모형에 기반한 이론과 기법들을 바탕으로 실제적인 통계자료처리에 적용하는 방법을 다룬다.

• **ECE741 컴퓨터비전 (Computer Visio)**

컴퓨터 비전의 일반적인 알고리즘들에 대해서 공부한다. 컴퓨터 비전이란 정지영상이나 동영상으로부터 3차원 환경을 분석하고 모델링하는 것을 목표로 한다. 이 과정에서는 먼저 영상의 기본적인 filtering, sampling 등의 개념을 배우고 edge detection, projection, image matching, motion estimation, image segmentation 등 컴퓨터 비전 각 분야의 대표적인 알고리즘들과 이들의 수학적 모델들에 대해 다룬다. 학기말에는 각자 컴퓨터 비전 최신 논문에서 제안하는 알고리즘을 구현하고 개선하는 프로젝트를 수행하면서 컴퓨터 비전의 노하우를 깨치게 된다.

• **ECE631 확률및스토캐스틱프로세스 (Probability and Stochastic Process)**

집합에 의한 확률이론과 대표적인 확률함수의 수리적 모형, 결합확률과 조건부확률의 개념, 랜덤변수의 개념 및 분포함수, 밀도함수, 기대값, 모멘트와 상관의 개념을 공부하고, 랜덤변수의 함수에 대한 확률적 연산을 익힘으로 관련된 공학분야에의 적용과 랜덤 신호 및 스토캐스틱 프로세스 이론 등을 배우기 위한 기초를 얻게 한다.

• **MED993 의학통계학 (Statistics in Medical Research)**

의학연구를 수행하는데 필요한 통계학적 개념과 통계방법론의 기본적인 지식을 강의하고, 연구설계와 연계한 적용방법을 토의함으로써, 실제 의학연구에 있어서의 응용력과 바른 통계분석법에 대한 판단력을 습득하도록 한다.

• **CSE6110 고급데이터마이닝 (Advanced Data Mining)**

데이터 마이닝과 기계 학습 분야의 고급 수준 강좌로, 고전적인 방법론에서부터 최신 학습 알고리즘에 이르기까지 실제 응용에서 유용하게 사용되고 있는 다양한 기법들을 소개한다. 다양한 분류 (classification) 기법, 고차원 회귀분석 (regression) 모델, 군집화 (clustering), bagging and boosting, 요인 분석 (factor analysis), 은닉마르코프 모델 (hidden markov model), 그리고 확률 그래프 모델 (probabilistic graphical model) 등을 다룬다.

• **DS751 의료정보분석특론 (Advanced medical data analysis)**

의료 분야에서 요구되는 데이터 분석 방법과 고려 사항들에 대해 배우기 위해, 실제 사례에 대한 발표와 토론을 중심으로 진행 된다. 본 과정은 향후 수강생들이 의료 분야 데이터 분석 현장에서 주어진 과제에 대해 능동적으로 대처할 수 있는 능력을 키울 수 있도록 한다.

• **MATH681 수학적 모델링 I (Mathematical Modeling I)**

물리과학(Physical Sciences), 공학 등에 등장하는 과학적 현상을 수리 모델링 하는 방법을 배운다. 구체적으로, 실제 문제에서 등장하는 미분방정식, 선형 시스템, 비선형시스템, 알고리즘 등을 학습한다.

• **DS722 기계학습특론 (Advanced Topics in Machine Learning)**

기계 학습과 관련된 최신의 이론 및 응용, 추세에 대하여 다룬다.

MATH671 산업수학 특론 I (Industrial Mathematics Survey I)

수학이 요구되는 구체적인 산업 문제와 그 해결을 위해 적용가능한 수학적 도구들을 조사하고, 팀 프로젝트 수행 시 필요한 보고서 작성과 구두 발표를 연습한다.

MATH672 산업수학 특론 II (Industrial Mathematics Survey II)

수학이 요구되는 구체적인 산업 문제와 그 해결을 위해 적용가능한 수학적 도구들을 조사하고, 팀 프로젝트 수행 시 필요한 보고서 작성과 구두 발표를 연습한다.

• **DS560~DS562 현장실습1~3 (Field Practice 1~3)**

현장체험을 통하여 전공 지식을 활용하고 학습 동기를 부여 하는 것을 목표로 한다. 학생은 현장실습을 할 회사를 정하고 회사의 프로젝트에 참여하여 학교에서 배운 기초 지식을 응용하는 현장실습의 기회를 가지게 되며, 회사에 근무하면서 회사의 실무자와 같이 현장의 문제해결방법을 체득한다.