

326.521: 고급통계적 방법론

Syllabus, 2025 가을학기

Course Home Page: eTL

이 과목에서는 컴퓨팅의 발전을 통해서 통계학분야가 지난 반세기동안 어떻게 발전했는지 알아본다. 특히 빅데이터의 시대를 맞이하여 탐색적 자료분석과 확증적 자료분석을 보다 일반화한 개념인 알고리즘과 통계적 추론의 발전과정에 대해서 공부한다. 알고리즘은 기본적으로 “어떻게 분석을 하는냐”를 초점을 맞추고 있으며 이러한 경향은 특히 데이터 사이언스로 대표되는 분야에서 특히 강조되고 있다, 반대로 통계적 추론은 “왜 이렇게 분석을 해야 하는냐”에 대한 답변을 제공하기 위한 수학적 논리를 제공하는 분야라고 할 수 있다. 이 과목에서는 통계학의 주요 방법론을 연대별로 다음과 같이 크게 3단계로 나누어 다룰 예정이다.

Part I 컴퓨팅 도약 이전 시점(1950년대)을 중심으로 통계학 분야의 대표적인 3개의 학파, 베이저안(Bayesian), 빈도주의(Frequentist), 우도주의 (Fisherian)에 대해서 알아본다.

Part II 컴퓨터의 초창기 도입시기인 1950년대부터 1990년대까지 개발된 대표적인 통계 방법론, Resampling methods, 생존분석과 EM 알고리즘, 일반화선형모형, Empirical Bayes, MCMC에 대해 소개한다.

Part III 21세기 빅데이터의 시대에 등장한 대표적인 통계방법론, 다중비교와 벌점화 회귀분석, Random Forests, Neural Network, SVM, Post selection inference를 대해 공부한다.

담당교수와 강의조교

담당교수:

장원철 25동 323호, 880-8164 wcjang@snu.ac.kr

강의조교:

정우진 25동 307호 cowzin@snu.ac.kr

한중건 25동 307호 hih1210@snu.ac.kr

강의시간과 Office Hours

강의 월수 오후 2시 - 오후 3시 15분 25동 210호

Office Hours

장원철 수 오후 4-6시 25동 323호

정우진 목 오후 2-3시 25동 307호

한종건 금 오후 2-3시 25동 307호

선수과목

다음 과목을 선수 과목으로 한다.

- 통계학과 학부교과목 *수리통계 1, 2* (326.311, 326.312)
- 통계학과 학부교과목 *데이터마이닝 방법 및 실습* (326.413) 또는 컴퓨터공학부 학부교과목 *기계학습개론* (4190.420)

즉 선형대수학 기초지식과 참고문헌 Wasserman (2004)에 소개되는 통계이론을 알고 있는 것을 전제로 한다. 만약 위의 선수과목을 수강하지 않은 경우 담당교수와 개별면담을 통해 수강신청여부를 결정한다. 다음과 같은 내용을 반드시 숙지해야 한다.

1. Convergence in probability and distribution.
2. Maximum likelihood estimation: Fisher information
3. Hypothesis testing
4. Bayesian inference
5. Linear regression
6. Logistic regression
7. Regularization
8. Bayes classifiers.

9. Support vector machine
10. Determinants, eigenvalues and eigenvectors

교재와 참고문헌

강의는 Efron and Hastie (2021)를 기반으로 진행될 예정이다.

- Efron, B. and Hastie, T. (2021). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, Student Edition. Cambridge University Press. ISBN 978-1108823418 (교재) - Available at <https://hastie.su.domains/CASI/index.html>.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer. ISBN 978-0387848570. (참고문헌) - Available at <https://web.stanford.edu/~hastie/ElemStatLearn/>.
- Murphy, P. K. (2022). *Probabilistic Machine Learning: An Introduction*. ISBN 978-0262046824. (참고문헌) - Available at <https://probml.github.io/pml-book/book1.html>.
- Wasserman, L. (2004). *All of Statistics: Concise Course in Statistical Inference*. Springer. ISBN 978-0387402727. (참고문헌) - Available at <http://link.springer.com/book/10.1007/978-0-387-21736-9> on campus

평가방법

- 격주로 숙제가 나갈 예정이며 숙제는 숙제제출 마감날짜 오후 5시까지 25동 2층의 숙제 제출함에 제출하여야 한다. 늦게 제출한 숙제는 0점 처리할 예정임.
- 시험일정. 중간고사와 기말고사 일정은 다음과 같다.
 - 중간고사: 10월 24일 오후 6-8시 (금요일)
 - 기말고사: 12월 12일 오후 6-9시 (금요일)

최종학점은 다음과 같이 계산된다.

- 수업참여도/출석 5%
- 숙제 35%
- 중간고사 30%
- 기말고사 30%

프로그래밍 언어

이 과목에서 사용되는 예제는 모두 R를 사용하여 분석결과를 제시할 예정이다. R은 통계분석에 최적화된 언어로 파이썬, Matlab과 비교시 자료분석에 있어서 여러가지 장점이 있다. R에 관한 다양한 튜토리얼과 참고문헌은 <http://www.r-project.org>에서 찾아볼 수 있으며 R에 관한 (조금 오래된) 뉴욕타임즈의 기사는 <https://www.nytimes.com/2009/01/07/technology/business-computing/07program.html>에서 읽을 수 있다. 한국어로 된 참고문헌으로는 *R for Practical Data Analysis* (<http://r4pda.co.kr/>)이나 권재명 (2017) 따라하면서 배우는 데이터 과학 (<http://dataninja.me/ipds-kr/>)을 추천한다.

숙제협력에 관한 원칙

숙제를 수업을 듣는 동료학생들과 같이 협력하는 것을 장려한다. 하지만 답안지와 분석에 사용된 코드는 반드시 본인 스스로 작성해야 되며 도움을 받을시에는 도움을 준 학생의 이름을 숙제제출시 명시해야한다. 그렇지 않을 경우 같은 답안을 제출한 숙제는 모두 0점 처리한다.

수업시간에 준수해야 할 사항

- 수업시간에 다른 사람과 대화 등으로 수업을 방해하지 않는다.
- 수업에 늦을 경우 다른 학생들에게 방해되지 않게 조용하게 들어와 앉는다.
- 전자기기(휴대폰, 아이패드, 컴퓨터 등)는 필기용으로 사용하는 경우에 **앞에서 3번줄까 지 한해서** 수업시간에 사용이 허용된다.
- 특별한 이유없이 결석을 잦을 경우 수업참여 점수가 0점처리된다.

장애학생에 대한 지원서비스 안내

장애유형	지원서비스	
	강의수강관련	과제 및 평가관련
시각장애	교재재작 (디지털교재, 점자교재, 확대교재등) 대필도우미 허용	과제제출기한 연장 과제제출방식 및 응답방식의 조정 평가시간 연장 평가 문항제시 및 응답방식의 조정 별도 고사실 제공
지체장애	교재 제작 (디지털교재) 대필도우미 및 수업보조도우미 허용	
청각장애	대필및 문자통역 도우미 활동허용 강의 녹취허용	
건강장애	질병등으로 인한 결석에 대한 출석 인정 대필도우미 허용	
학습장애	대필도우미 허용	
지적장애 자폐성장애	대필도우미 및 수업멘토 허용	개별화 과제 제출및 대체평가 실시

Schedule

Date	Topic	Reading	HW due
9월 1일	Introduction	Handout & Ch 1	
9월 3일	Frequentist Inference	Ch 2	
9월 8일	Bayesian Inference I	Ch 3	
9월 10일	Fisherian Inference II	Ch 4	HW 1 due (9/12)
9월 15일	Exponential Families I	Ch 5	
9월 17일	Exponential Families II	Ch 5	
9월 22일	Empirical Bayes I	Ch 6	
9월 24일	Empirical Bayes II	Ch 6	HW 2 due (9/26)
9월 29일	Shrinkage Estimation I	Ch 7	
10월 1일	Shrinkage Estimation II	Ch 7	
10월 6일	추석연휴 (공휴일)		
10월 8일	추석연휴 (공휴일)		
10월 13일	Shrinkage Estimation III	Ch 7	
10월 15일	개교기념일 (정기휴업일)		HW 3 due (10/17)
10월 20일	GLM I	Ch 8	
10월 22일	GLM II	Ch 8	
10월 24일	중간고사		
10월 27일	Survival Analysis I	Ch 9	
10월 29 일	Survival Analysis II	Ch 9	HW 4 due (10/31)
11월 3일	Resampling I	Ch 10	
11월 5일	Resampling II	Ch 10	
11월 10일	Bootstrap Confidence Intervals I	Ch 11	
11월 12일	Bootstrap Confidence Intervals II	Ch 11	HW 5 due (11/14)
11월 17일	Cross-Validation I	Ch 12	
11월 19일	Cross-Validation II	Ch 12	
11월 24일	Multiple testing I	Ch 15	
11월 26일	Multiple testing II	Ch 15	HW 6 due (11/28)
12월 1일	Empirical Bayes Estimation Strategies I	Ch 21	
12월 3일	Empirical Bayes Estimation Strategies II	Ch 21	
12월 8일	Modern Dependency Measure	Lecture notes	
12월 10일	Time Series Forecasting	Lecture notes	HW 7 due (12/12)
12월 12일	기말고사		