전북대학교
카피킬러캠퍼스 표절 검사
결과 확인서

확 인
성 명
서 명

아이디	202055364	표절률	7%
소속	대학원 통계학과		
성명	자필로 기재하세요		

검사번호	00151892239	검사일자	2021.12.03 00:22
발급형태	▫ 기본보기 ▫ 요약보기 ■ 상세보기	발급일자	2021.12.03 00:24
검사명	중간점검(학위논문 표절검사)		
문서명	박사학위논문_작성중_황성윤.hwp		
비고			

비교범위	[현재첨부문서] [카피킬러 DB]
검사설정	표절기준 [6 어절], 인용/출처 표시문장 [제외], 법령/경전 포함문장 [제외], 목차/참고문헌 [제외]

검토 의견	

분석 정보

표절률	전체문장	동일문장	의심문장	인용/출처	법령/경전
7%	367	0	86	0	0

비교 문서 정보

번호	표절률	출처정보	비고
1	3%	[카피킬러 DB] Copykiller - 파일명 : 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method - 저자 : 黃星潤 - 발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2	
2	2%	[카피킬러 DB] Copykiller - 파일명 : 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method - 저자 : 황성윤 - 발행 : 서울 : 한국외국어대학교 대학원, 2017.2	
3	1%	[카피킬러 DB] Copykiller - 파일명 : IS - 저자 : Harry H. Kelejian, Ingmar R. Prucha, Yevgeny Yuzefovich	
4	1%	[카피킬러 DB] Copykiller - 파일명 : [Advances in Econometrics] Spatial and Spatiotemporal Econometrics Volume 18 || INSTRUMENTAL VARIABLE ESTIMATION OF A SPATIAL AUTOREGRESSIVE MODEL WITH AUTOREGRESSIVE DISTURBANCES: LARGE AND SMALL SAMPLE RESULTS - 저자 : Kelejian, Harry H. - 발행 : 2004	
5	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture (14:00 ... - 발행 : radioactivity.nsr.go.jp	
6	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture (14:00 ... - 발행 : radioactivity.nsr.go.jp	
7	1%	[카피킬러 DB] drum.lib.umd.edu - 파일명 : ABSTRACT Title of Dissertation: TWO ESSAYS ON SPATIAL ···	
8	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture (14:00 ... - 발행 : radioactivity.nsr.go.jp	
9	1%	[카피킬러 DB] Copykiller - 파일명 : Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids. - 저자 : Iffat Parveen, Hemant Kumar Singh, Saloni Malik, Saurabh Raghuvanshi, Shashi B Babbar - 발행 : 2017	
10	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture (14:00 ... - 발행 : radioactivity.nsr.go.jp	
11	1%	[카피킬러 DB] Copykiller - 파일명 : ABSTRACT Title of Dissertation: TWO ESSAYS ON SPATIAL ECONOMETRICS - 저자 : Yevgeniy A Yuzefovich, Doctor Of Philosophy, Dissertation Professors, Harry Kelejian, Ingmar Prucha	
12	1%	[카피킬러 DB] Copykiller - 파일명 : Evaluating five different Loci (rbcL, rpoB, rpoC1, matK and ITS) for DNA Barcoding of Indian Orchids - 저자 : Parveen, Iffat, Singh, Hemant K., Malik, Saloni, Raghuvanshi, Saurabh, Babbar, Shashi B. - 발행 : 2017-04-21	

13	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture（14:00 ... - 발행 : radioactivity.nsr.go.jp
14	1%	[카피킬러 DB] usermanual.wiki - 파일명 : Poisson Superfish Los Alamos Manual
15	1%	[카피킬러 DB] Copykiller - 파일명 : UTAUT2 Based Predictions of Factors Influencing the Technology Acceptance of Phablets by DNP - 저자 : Huang, Chi-Yo; Kao, Yu-Sheng - 발행 : 2015
16	1%	[카피킬러 DB] Copykiller - 발행 : 2013.07.
17	1%	[카피킬러 DB] www.boxnwhis.kr - 파일명 : 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker
18	1%	[카피킬러 DB] boxnwhis.kr - 파일명 : 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker - 발행 : boxnwhis.kr
19	1%	[카피킬러 DB] hoon427.tistory.com - 파일명 : 서포트 벡터 머신 - 발행 : hoon427.tistory.com
20	1%	[카피킬러 DB] Copykiller - 파일명 : GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the Brand Image Positioning on PCA, Clustering and G-LDA Topic Modeling Using GloVe - 저자 : 박영욱 - 발행 : 2021
21	1%	[카피킬러 DB] Copykiller - 파일명 : GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the Brand Image Positioning on PCA, Clustering and G-LDA Topic Modeling Using GloVe - 저자 : 박영욱 - 발행 : 2021
22	1%	[카피킬러 DB] Copykiller - 파일명 : GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : ... - 저자 : 박영욱 GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : the case of domestic 5 stars hotel reviews on Trip Advisor web site
23	1%	[카피킬러 DB] Copykiller - 파일명 : 일가족양립제도가 기혼여성근로자의 취업중단에 미치는 영향 - 저자 : 임지영 - 발행 : 서울 : 연세대학교 사회복지대학원, 2011
24	1%	[카피킬러 DB] Copykiller - 파일명 : 취업지원서비스의 여성 고용유지 효과 = Women's Vocational Training & Reentry to Job Market - 저자 : 오은진, 김소연 - 발행 : 서울 : 한국직업능력개발원, 2017.11.30
25	1%	[카피킬러 DB] Copykiller - 파일명 : 교차하는 두 생존함수의 동일성 검정법에 관한 비교연구 - 저자 : Journal of the Korean Data & Information Science Society = 한국데이터정보과학회지 v.26 no.3 , 2015년, pp.569 - 580 이윤주 (고려대학교 통계학과) ; 이재원 (고려대학교 통계학과) - 발행 : 2015
26	1%	[카피킬러 DB] Copykiller - 파일명 : 취업지원서비스의 여성 고용유지 효과 - 발행 : 2017 vol.20 no.3 pp.149-178
27	1%	[카피킬러 DB] Copykiller - 파일명 : Chapter 12.B Skilled and Unskilled Labor Data - 저자 : Betina V. Dimaranan, Badri Narayanan G

28	1%	[카피킬러 DB] radioactivity.nsr.go.jp - 파일명 : Reading of environmental radioactivity level by prefecture（14:00 ... - 발행 : radioactivity.nsr.go.jp
29	1%	[카피킬러 DB] Copykiller - 파일명 : 연구보고서 20-14 조세 · 재정 정책과 기업의 고용조정에 관한 연구: 청년고용 임금보조금과 세액공제를 중심으로 - 저자 : 김문정 오종현 조원기 - 발행 : 2021-03-31
30	1%	[카피킬러 DB] Copykiller - 파일명 : Current profile and sea-bed pressure and temperature records from the northern North Sea. Challenger Cruises 84 and 85. September 1991 - November 1991 - 저자 : Knight, P. J.Wilkinson, M.Glorioso, P. - 발행 : 1993

검사 문서	비교 문서

표지

博士學位論文 중도절단이 포함된 생존자료에 대한 회귀분석에서 커널트릭 기법과 앙상블 기법을 적용했을 경우의 예측력 향상에 관한 연구 A study on the improvement of predictive power when the kernel trick method and the ensemble method are applied in the regression analysis of survival data including censoring 指導 梁 城 準 敎授 2024年 2月 17日 全北大學校 大學院 統 計 學 科 黃 星 潤 博士學位論文 중도절단이 포함된 생존자료에 대한 회귀분석에서 커널트릭 기법과 앙상블 기법을 적용했을 경우의 예측력 향상에 관한 연구 A study on the improvement of predictive power when the kernel trick method and the ensemble method are applied in the regression analysis of survival data including censoring 指導 梁 城 準 敎授 이 論文을 博士學位請求論文으로 提出합니다. 2023年 11月 全北大學校 大學院 統 計 學 科 黃 星 潤 이 論文을 黃星潤의 博士學位論文으로 認定함. 委 員 長 전북대학교 교수 (인) 副委員長 전북대학교 교수 (인) 委 員 전북대학교 조교수 (인) 委 員 전북대학교 조교수 (인) 委 員 전북대학교 조교수 (인)

문장표절률: **0%**

2023年 月 日 全北大學校 大學院

목차

문장표절률: **0%**

ABSTRACT A study on the improvement of predictive power when the kernel trick method and the ensemble method are applied in the regression analysis of survival data including censoring SEONG YUN HWANG DEPARTMENT OF STATISTICS THE GRADUATE SCHOOL

문장표절률: **0%**

JEONBUK NATIONAL UNIVERSITY This study relatesto a method that can improve predictive power when regression analysis is performed on data with censoring.

문장표절률: **0%**

Censoring is usually caused by internal or external causes such as the patient's death due to factors other than the disease being studied in the survival data related to the patient's survival time, which appears frequently in themedical field.

문장표절률: **0%**

The main purpose of analyzing survival data is to determine which factors have a significant effect on the patient's survival time and to predict the patient's survival time through this.

문장표절률: 0%

In the case of survival data including such censoring, since the survival time, wh ich is the subject of estimation, is only partially observed, it is possible to analyze the data by creating an synthetic response to replace it.

문장표절률: 0%

However, these synthetic response have a characteristic that the conditional vari ance when explanatory variables are given tends to be larger than the conditional variance of the originalsurvival time, and the width increases as the survival time increases.

문장표절률: 0%

Because of this, thestability of the estimator is poor, which can be a problem. In order to compensate for this problem, in this study, when constructing a regress ion model for synthetic response, kernel trick method and ridge regression meth od are applied.

문장표절률: 0%

Kernel trick method often used when data in the explanatory variablespace is mo ved to a high-dimensional characteristicspace by using an appropriate mapping function for complex nonlinear data without specifying a transformation functio nseparately.

문장표절률: 0%

And ridge regression method applicable when there is a problem of multicollinear ity. In addition, we would like to propose a method for improving the predictive power of survival time by reducing the variance of the estimator by applying ens emble techniques such as bagging and random forest.

문장표절률: 0%

Through computersimulation, various situations were assumed and the predictiv e power of explanatory variables was compared and analyzed in the data includi ng censoring.

문장표절률: 45%

Through this, it was confirmed that the method proposed in this study showed o verall superior predictive power compared to the general method.

[Copykiller] 머신 러닝을 이용한 풀필먼트 센터의 출고 운영의 위험관리 = Risk Ma nagement in Outbound Operation of Fulfillment Center using Machine Learnin g

저자 : 김창현
발행 : 학위논문(박사)-- 인천대학교 동북아물류대학원 : 물류시스템학과 2020. 2

can also be relatively reduced.It was confirmed that the method proposed in this study could improve the situation of

[Copykiller] 머신 러닝을 이용한 풀필먼트 센터의 출고 운영의 위험관리 = Risk ma nagement in outbound operation of fulfillment center using machine learning (다운로드)

저자 : 김창현 머신 러닝을 이용한 풀필먼트 센터의 출고 운영의 위험관리 = Risk management in outbound operation of fulfillment center using machine learning

can also be relatively reduced. It was confirmed that the method proposed in thi s study could improve the situation of

문장표절률: 59%

It is expected that the method proposed in this study can be used efficiently whe n analyzing datacontaining censoring from various research fields.

[Copykiller] A study on large-area laser processing analysis in consideration of the moving heat source

저자 : Sung-Hwan Ahn; Choon-Man Lee
발행 : 2011-04

routes will reduce the error. It is expected that the method proposed in this study can be used as a simple analysis method

[Copykiller] A Study on Large-area Laser Processing Analysis in Considerati on of the Moving Heat Source

저자 : Ahn, Sung-Hwan,Lee, Choon-Man
발행 : 2011

routes will reduce the error. It is expected that the method proposed in this study can be used as a simple analysis method

문장표절률: 0%

keywords : survival data, survival analysis, survival time, censoring, synthetic response, ridge regression, machine learning, kernel trick method, ensemble method 1 서 론 1.1 연구의 배경 및 목적 생존자료(survival data)는 주로 의학 분야에서 환자의 생존시간(survival time)을 확인하고 연구할 때 사용하는 데이터이다.

문장표절률: 0%

최근에는 의학 분야 외에도 경제, 경영분야에서 기업의 생존율 또는 노동자의 실업률을 분석하는 등의 다양한 연구 분야에서 폭넓게 사용되고 있으며 그 활용도 또한 증가하고 있는 추세이다.

문장표절률: 0%

특별히 생존자료의 가장 큰 특징 중의 하나는 환자가 연구대상인 질병 이외의 원인에 의해 사망하는 경우, 입원해 있던 환자가 다른 병원으로이송되는 경우, 연구자가 임의로 관측하는 시간을 조정하는 경우, 진행하던 연구가 중단되는 경우 등의 내부적 또는 외부적인 요인에 의해 발생하는 중도절단(censoring)이 포함되어 있다는 것이다.

문장표절률: 0%

그러므로 생존자료에서는 추정의 대상이 되는 실제 환자의 생존시간(survival time)이 중도절단이 일어나지 않은 경우에 한정하여 부분적으로만 관측된다.

문장표절률: 0%

이러한 생존자료의 특징을 바탕으로 카플란-마이어 추정법(Kaplan-Meier estimation), 넬슨-알렌 추정법(Nelson-Aalen estimation), 콕스 비례위험모형(Cox proportional hazard model), 모수적 모형(Parametric model), 가속실패시간모형(Accelerated failure time model) 등 다양한 분석 방법이 제안되었다.

문장표절률: 0%

다양한 생존 분석방법에 대한 자세한 사항은 Kleinbaum, D.G. and Klein, M. (2010)에서 확인할 수 있다.

문장표절률: 0%

하지만 최근 들어 빅데이터(big-data) 시대에 우리가 실질적으로 마주하게 되는 데이터는 반응변수(response variable)와 설명변수(explanatory variable)와의 관계가 비선형(nonlinear relationship)인 경우이거나 정형화(standardization)가 되어있지 않은 경우가 대부분이기 때문에 분석시 변수변환(variable transformation) 등의 다양한 방법을 고려하여 분석을 진행하기에 적절한 형태를 가지도록 연구자가 미리 데이터를 처리해 주어야 하는 경우가 많이 발생한다.

문장표절률: 0%

또한 데이터에 포함되어 있는 설명변수 사이에 있는 연관성 때문에 발생하는 다중공선성(multicollinearity)에 대한 문제도 극복해야할 중요한 사항 중 하나이다.

문장표절률: 0%

이러한 문제점은 의학 분야에서 다루는 생존자료(survival data)에서도 예외가 아니다. 따라서 본 논문에서는 첫 번째 연구결과로서 중도절단이 포함된 생존자료를 분석하는 경우 이와 같은 문제점을 어느 정도 극복할 수 있게 해주는 방법을 제안한다.

문장표절률: 0%

두 번째 연구결과 설명... 1.2 연구방법 및 구성 본 논문에서는 크게 2가지의 연구를 수행하였으며 모두 생존자료에 대한 분석 시 기계학습방법을 적용하는 방안에 관한 내용을 담고 있다.

문장표절률: 0%

먼저 첫 번째 연구에 대한 내용을 전체적으로 설명하도록 하겠다. 중도절단이 포함된 생존자료에서 생존시간을 예측하고자 할 경우 부분적으로만 관측되는 실제 환자의 생존시간을 대체하기 위하여 정의된 인조변수(synthetic response)를 모형을 구축할 때 사용한다.

문장표절률: 0%

이러한 인조변수를 이용한 자료변환법에 대해서는 Buckley, J. and James, I. (1979), Koul et al. (1981), Leurgans, S.

문장표절률: **17%**

(1987) 등을 참조하기 바란다. 그리고 이 인조변수를 반응변수로 두고 회귀분석(regression analysis)을 실시하되 커널트릭 기법(kernel trick method)과 능형 회귀분석(ridge regression) 방법을 결합한 커널 능형 중도절단 회귀분석(kernel ridge censored regression) 방법을 적용한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

에 커 널기법 kernel trick)과 능형 회귀분석 ridge regression) 방법을 적용한 커널 능형 로지 스틱

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

classification)에 커널기법(kernel trick)과 능형 회귀분석(ridge regression) 방법을 적용한 커널 능형 로지 스틱

문장표절률: **0%**

커널트릭 기법은 사전에 분석하고자 하는데이터에 적합한 변환함수(transformation function)를 적용하지 않더라도 자동적으로 적절한 변환이 이루어지게 할 수 있는 방법으로 최근에 화두가 되고 있는 기계학습(machine learning) 분야에서 광범위하게 언급되고 있으며, 특히 분류(classification)문제에 자주 사용되는 지지도 벡터 기계 기법(SVM: support vector machine) 등에서 많이 활용되고 있다.

문장표절률: **0%**

그리고 능형 회귀분석(ridge regression) 방법은 다중공선성의 문제를 보완하기 위해 사용할수 있는 벌점회귀(penalized regression) 방법의 일종이다.

문장표절률: **0%**

여기에 추정량의 분산을 큰폭으로 줄일수 있는 앙상블 기법(ensemble method)을 추가로 적용하여 보다 안정적이고 신뢰성 있는 예측모형을 구축한다.

문장표절률: **10%**

본 연구에서는 반복적인 복원추출(sampling with replacement)을 통해서 뽑은 다수의 부트스트랩 표본(bootstrap sample)에 의한 결과를 평균하여 분산을 감소시키고 과대적합(overfitting)의 위험을 줄일 수 있는 배깅(bagging), 그리고 부트스트랩 표본추출 시 모든 설명변수를 포함시키지 않고 이들중 일부만을 선택하는 과정을 통해 설명변수 사이의 연관성을 줄임으로써 좀더 정확한 추정을 할 수 있도록 해주는 랜덤포레스트(random forest)를 적용한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

하나의 훈련자료 (training data) 에서 반복적인 복원추출 sampling with replacement)을 통하여 훈련자료와 비교했을 때 관측치의

문장표절률: **0%**

이러한 과정을 통하여 생존시간에 대한 예측력(predictive power)을 높이는 방법을 첫 번째 연구결과로 제안한다.

문장표절률: **0%**

다음으로는 두 번째 연구에 대한 내용을 전체적으로 소개하도록 하겠다. 두 번째 연구결과 요약 설명... 본 논문은 총 7장으로 구성되어 있다.

문장표절률: **0%**

1장에서는 본 연구의 배경과 이에 따른 연구방법을 개괄적으로 설명하고, 2장에서 6장까지는 중도절단이 포함된 생존자료에 대한 회귀분석시 커널트릭 기법과 앙상블 기법을 적용하는 경우의 예측력 향상에 대한 연구 내용을 설명한다.

문장표절률: **0%**

2장에서는 생존자료의 특성과 이를 분석할 때 일반적으로 사용되는 생존분석(survival analysis) 방법에 대하여 간략한 설명한 뒤 중도절단자료 분석을 위한 인조변수를 어떻게 정의할 수 있는지 언급한다.

문장표절률: **0%**

3장에서는 본 연구의 핵심인 커널 능형 중도절단 회귀분석에 관하여 설명하고, 4장에서는 앙상블 기법을 소개한다.

문장표절률: 0%

이어서 5장에서는 커널 능형 중도절단 회귀분석에 앙상블 기법을 어떻게 적용할 수 있는지 설명하며, 6장에서는 앙상블 기법을 이용한 커널 능형 중도절단 회귀분석 방법이 일반적인 방법보다 전반적으로 예측력이 우수하다는 사실을 입증하기 위해 실시한 모의실험과 실증분석의 결과를 제시하고 평가한다. 마지막으로 7장에서는 본 연구에 대한 결론을 언급하고 마무리한다.

문장표절률: 0%

2 생존자료 (survival data) 2.1 생존자료 (survival data) 생존자료(survival data)는 보통의학 분야에서 환자의 생존시간(survival time)을 분석하고 연구하기 위해 자주 사용된다.

문장표절률: 0%

서론에서도 언급했듯이 이 데이터의 가장 큰 특징은 환자가 연구대상인 질병 이외의 원인에 의해 사망하거나 연구자가 임의로 관측하는 시간을 조정하는 등의 내부적 또는 외부적인 요인에 의해 발생하는 중도절단(censoring)이 포함되어 있다는 것이다.

문장표절률: 0%

이 데이터에는 환자의 생존 시간에 영향을 줄 것이라고 판단하여 연구자가 포함시킨 설명변수 , 관측된 생존시간 , 그리고 중도절단의 여부를 나타내는 지시변수(indicator variable)인 가 기본적으로 포함되어 있다.

문장표절률: 0%

여기에서는 관심사건이 일어날 때까지 걸리는 시간(보통의학 분야에서 얘기하는 환자의 실제 생존시간), 는 중도절단된 시간을 의미한다.

문장표절률: 0%

즉, 를 만족하여 환자의 실제 생존시간이 관측된 경우에는 , 이 되며, 반대로를 만족하여 중도절단이 된 경우에는 , 이 되는 것이다.

문장표절률: 0%

다시 말해서 실제 생존자료에서는 모든 환자의 실제 생존시간을 확인할 수 없고 인 경우에 한하여 부분적으로만 실제 환자의 생존 시간을 관측할 수 있다.

문장표절률: 0%

그렇기에 이러한 생존자료를 분석하기 위한 특수한 방법이 필요한 것이고 이를 통틀어서 생존분석(survival analysis) 방법 이라고 부른다.

문장표절률: 0%

2.2 생존분석 (survival analysis) 생존분석에서는 기본적으로 다음과 같은 개념을 관측된 생존시간 를 바탕으로 정의한다.

문장표절률: 0%

여기에서는 생존함수(survival function), , 는 각각 위험함수(hazard function)와 누적위험함수(cumulative hazard function), 그리고 , 는 각각 생존시간에 대한 분포함수(distribution function)와 누적분포함수(cumulative distribution function)이다.

문장표절률: 0%

그리고 아래에 있는 〈그림 1〉은 프로그램 R 4.1.1 version을 이용하여 생존분석시 사용하는 패키지인'survival'에 내장되어 있는 데이터 중 하나인 폐암(lung cancer)에 걸린 환자와 관련한 'cancer' 데이터를 바탕으로 카플란-마이어 추정법(Kaplan-Meier estimation)을 적용하여 성별(sex)에 따라 생존함수를 추정하고 그 결과를 바탕으로 그린 추정된 생존함수의 그래프이다.

문장표절률: 0%

그리고 이 그래프 안에 표시된 유의확률은 성별에 따라 생존함수가 차이가 있는지에 대한 로그순위검정(log-rank test)의 결과를 나타낸 것이다.

문장표절률: 0%

결국 유의확률이 일반적인 유의수준보다 매우 작은 값이므로 성별에 따라 생존율이 달라진다고 해석할 수 있다.

문장표절률: 0%

추가로 생존함수 그래프 아래에 있는 수치는 시간에 따른 생존환자의 수를 나타낸 것이다.

문장표절률: 0%

〈그림 1 : Survival function of lung cancer data〉 〈그림 1〉을 통해서 알 수 있듯이 일반적으로 생존함수 는 최초에 에서부터 시작해서 시간이 지나 갈수록 점차 으로 단조감소(monotone decreasing)하는 형태를 보이게 된다.

문장표절률: 0%

즉, 시간이 지나 갈수록 환자의 생존율이 줄어든다는 것을 의미한다. 물론 치유개체를 추가적으로 고려하는 생존치유모형(survival cure model)처럼 특별한 원인에 의하여 시간이 지나도 생존하는 개체가 있는 상황을 고려하는 경우도 존재하지만 보통은 〈그림 1〉의 형태처럼 에서 시작하여 으로 수렴하는 형태의 생존함수를 사용한다.

문장표절률: 0%

생존치유모형에 대한 자세한 사항은 Sposto, R. (2002) 등을 참조하기 바란다. 그리고 위에 있는 식을 통해서도 알 수 있듯이 생존함수 , 위험함수 , 그리고 생존시간에 대한 분포함수 는 서로 연관되어 있다는 특징이 있다.

문장표절률: 0%

그렇기 때문에 생존자료를 통해 생존함수를 추정하게 된다면 이를 이용하여 위험함수와 생존시간에 대한 분포함수 도 자연스럽게 추정할 수 있다.

문장표절률: 0%

보통 생존분석에서는 〈표 1〉에 제시된 바와 같이 생존함수를 비모수적 방법(non-parametric method)으로 추정하는 카플란-마이어 추정법(Kaplan-Meier estimation), 누적위험함수를 비모수적 방법으로 추정하는 넬슨-알렌 추정법(Nelson-Aalen estimation), 비교하고자 하는 집단들 간의 위험(hazard)이 추적조사 기간 동안 일정하게 비례한다는 가정 하에 위험함수를 준모수적 방법(semi-parametric method)으로 모형화하는 콕스 비례위험모형(Cox proportional hazard model), 특정 분포를가정하고위험함수를 모수적 방법(parametric method)으로 모형화하는 모수적 모형(Parametric model) 등이 주로 사용되고 있다.

문장표절률: 0%

이외에도 다양한 생존분석방법들이 존재하며 이에 대한 자세한 설명은 Sabin, C. and Petrie, A. (2019)와 Chen et al. (2017) 등을 참조하기 바란다.

문장표절률: 0%

Kaplan-Meier estimation : 시점에서의 발생 사건 수 : 시점에서 위험에 놓인 개체수 Nelson-Aalen estimation : 시점에서의 발생 사건 수 : 시점에서 위험에 놓인 개체수 Cox proportional hazard model

문장표절률: 0%

Parametric model The Exponential Model The Weibull Model The Rayleigh Model The Gompertz Model The Lognormal Model : standard cumulative normal distribution 〈표 1 : 생존분석 방법의 대표적인 예〉 2.3 중도절단자료 분석을 위한 인조변수 (synthetic response)

문장표절률: 0%

2.1절에서 언급한 바와 같이 생존자료에 포함된 생존시간과 관련한 변수는 환자에 대해 관측된 생존시간인 뿐이기 때문에 모든 환자에 대한 정확한 생존시간, 즉 관심사건이 일어날 때까지 걸리는 시간을 알 수 없다.

문장표절률: 0%

그러므로 생존시간을 예측하기 위한 회귀모형을 구축하는 경우에 관측된 생존시간 를 반응변수(response variable)로 둔다는 것은 약간의 무리가 있다.

문장표절률: 21%

따라서 본 연구에서는 Koul et al. (1981)에서 제안한다음과 같은 인조변수(synthetic response) 를 사용하여 자료를 변환하였다. 자세한 내용은 Kim, J. (2018)와 Lee, S. (2018)를 참조하기 바란다.

[Copykiller] 유전체 자료를 분석하기 위한 준모수 가변계수모형 연구

발행 : 2020

보였고 를 이용한 선형회귀모형을 제안하였다. 본 연구에서는 Koul et al. 1981)이 제안한 가중값( )을 이용하여

문장표절률: **22%**

, , , 여기에서 함수의 경우는 절단변수가 설명변수에 의존하지 않는다는 가정 하에 카플란-마이어 추정량(Kaplan-Meier estimator)을 사용하여 추정하게 되며, 만약의존성이 있는 경우에는 Beran, R. (1981)에서 제안한다음과 같은 형태의 추정량을 사용하여야 한다.

[boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

발행 : boxnwhis.kr

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다. 이론적인 생존 곡선과는 다르게

[www.boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다.

문장표절률: **0%**

단, 는 나다라야-왓슨 가중치(Nadaraya-Watson weight), 는 Bandwidth, 는 커널함수(kernel function)를 의미한다.

문장표절률: **0%**

본 연구에서는 절단변수가 설명변수에 의존하지 않는다는 가정을 바탕으로 다음과 같은 형태의 카플란-마이어 추정량을 사용하였다.

문장표절률: **0%**

한가지 주의해야할 점은 이러한 인조변수를 생성할 경우 분모에 대한 추정량의 값이거의 에 가깝게 되는 시점, 즉 인 경우에는 의 값이 무한대로 발산하거나 정의되지 않는다는 것이다.

문장표절률: **0%**

본 연구에서는 이러한 상황이 발생하는 경우를 대비하여 절단점(truncation point)을 정해주는 방법을 사용하였으며, 관측된 생존시간의 값이 누적확률이 이 되는 시점보다 더 큰 경우에는 으로 설정하였다.

문장표절률: **0%**

인조변수는 몇가지 적절한 조건을가 정한다고 할 때 다음과 같은 성질을 가지고 있다.

문장표절률: **0%**

여기에서 몇가지 적절한 조건은 우선 환자의 실제 생존시간 와 중도절단된 시간 는 서로 독립이고(), 가보다 작거나 같을 확률이 설명변수에 의존하지 않는다는 것()이다.

문장표절률: **0%**

---------- If , and , then proof) 1) If , then . So . 2) If , then . So . *---------- 즉, 관심사건이 일어날 때까지 걸리는 시간 와 중도절단된 시간 가 서로 독립이고 중도절단이 일어나지 않을 확률이 설명변수에 의존하지 않는다면 인조변수의 조건부평균(conditional mean)이 실제 관심사건이 일어날 때까지 걸리는 시간의 조건부평균과 일치하게된다.

문장표절률: **0%**

그렇기 때문에 변수를 반응변수로 두는 대신 인조변수를 새로운 반응변수로 두고 회귀모형을 구축하는 것이 타당하다고 할 수 있다.

문장표절률: **0%**

하지만 조건부분산(conditional variance)의 경우는 다음과 같이 인조변수의 경우가 변수의 경우보다 더 크다.

문장표절률: **0%**

---------- If , and , then proof) *---------- 이러한 성질 때문에 인조변수를 사용하여 관심사건이 일어날 때까지 걸리는 시간 를 예측하기 위한 회귀모형을 구축하게 되면 추정량에 대한 분산이 증가함에 따라 변동성도 커지게 되어 모형의 안정성과 신뢰성이 떨어지는 단점이 있다.

문장표절률: **0%**

이러한 문제점을 보완하기 위해 능형 회귀분석(ridge regression) 방법과 앙상블 기법(ensemble method)을 적용할 수 있으며 이에 대해서는 각각 3장과 4장을 통해서 소개하도록 하겠다.

문장표절률: 0%

3 커널 능형 중도절단 회귀분석 본 장에서는 커널 능형 중도절단 회귀분석방법에 관하여 설명한다.

문장표절률: 0%

이 방법은 기본적으로 생존자료를 통해 환자에게서 관심사건이 일어날 때까지 걸리는 시간을 예측하고 자하는 것이 주된 목적이며, 2.3절에서 설명한 인조변수(synthetic response)를 반응변수로 두고 다중회귀모형을 적합하게 된다.

문장표절률: 10%

여기에 회귀계수를 추정하는 경우 벌점함수(penalty function)를 통한 규제(regulation)를 바탕으로 다중공선성의 문제(multicollinearity problem)를 보완할 수 있는 능형 회귀분석(ridge regression) 방법과 반응변수와 설명변수 사이에 비선형 관계(non-linear relationship)가 존재하는 경우 이를 공간변환(space transformation)을 통해 적절히 해결해줄 수 있는 커널트릭 기법(kernel trick method)을 추가로 적용하여 보다 더 강한 예측력을 가지는 모형을 구축하는 방법이 바로 커널 능형 중도절단 회귀분석 방법이다.

[Copykiller] 커널능형회귀분석에서 앙상블기법을 이용한 효율성 연구 = A study on kernel ridge regression using ensemble method

저자 : 韓善雨
발행 : 韓國外國語大學校 大學院 : 통계학과 2016. 2

최근 비선형구조를 선 형구조로 변환시킬 수 있는 커널트릭 기법(kernel trick method)이 많은 관심 을 받고

문장표절률: 0%

이를 통해 생존자료를 분석하고 모형을 적합하여 관심사건이 일어날 때까지 걸리는 시간을 예측한다면 일반적인 방법을 통해 예측하는 것보다 더 높은 정확성을 바탕으로 좋은 예측결과를 얻을 수 있게 된다.

문장표절률: 0%

본격적으로 커널 능형 중도절단 회귀분석 방법에 대해 소개하기 전에 우선 3.1절을 통해 이 방법의 근간이 되는 다중회귀분석(multiple regression analysis)에 대해 설명 하도록 하겠다.

문장표절률: 21%

3.1 다중회귀분석 (multiple regression analysis) 다중회귀분석(multiple regression analysis)은 보통 하나의 반응변수(response variable)와 2가지 이상의 설명변수(explanatory variable)들 사이에 특정한 관계가 있다고 할 때 그 관계의 형태가 선형(linear)이라는 가정을 바탕으로 자료를 분석하고자 제안된 통계적인 분석방법이다.

[patents.google.com] KR100637939B1 - 고객 만족 지수 분석 시스템 및 방법 - Google ...

발행 : patents.google.com

관계가 거의 없다는 것을 의미한다. 다중회귀분석(Multiple Regression Analysis) 다중회귀분석(Multiple Regression Analysis)은 일반적으로 두 변수 이상의 독립변수

[Copykiller] 특발성 척추 측만증 환자의 모니터링을 위한 cobb-angle 예측 다중 회귀 모델 개발

저자 : 서은지
발행 : 서울 : 성균관대학교 대학원, 2012.8

경우 사용된다. - 16 - 3-4 다중회귀분석 (Multiple Regression Analysis) 다중회귀분석(Multiple Regression Analysis)은 둘 이상의 독립변수가 하나의 종

문장표절률: 0%

이에 대한 자세한 사항은 Han, S. (2016)과 Hwang, S. (2017)을 참조하기 바란다. 이해를 돕기위해 예시를 살펴보도록 하자.

문장표절률: 0%

만약 개의 관측치, 그리고개의 설명변수가 포함된 데이터 이 있다고 한다면 다중회귀분석에서는 다음과 같은 모형식을 기본적으로 가정하고 자료를 분석한다.

문장표절률: 0%

이 모형식을 통해서 알 수 있듯이 다중회귀분석 방법은 반응변수와 설명변수들 사이에 선형적인 관계가 존재한다고 기본적으로 가정한다.

문장표절률: 0%

여기에서는 다중회귀모형에서 가정하는 오차(error)를 뜻하며 일반적으로 모두 평균이, 그리고 분산이 인 정규분포(normal distribution)를 따르고 서로 독립(iid : identical and independently distributed)이라는 가정을 하게된다.

문장표절률: 0%

위에 있는 회귀모형식은 보통 다음과 같은 행렬(matrix)과 벡터(vector)의 형태로 표현하는 것이 일반적이고 이러한 형태의 표현식이 이론적인 증명을 위한 계산적인 측면이나 가독성의 측면에서는 더 좋은 표현 방법 이라고 필자는 생각한다.

문장표절률: 0%

여기에서는 길이가 인 반응변수벡터(response variable vector), 는 길이가 인 오차벡터(error vector)를 의미한다. 그리고 는 길이가 인 회귀계수벡터(regression parameter vector)이고

문장표절률: 0%

는 크기가 인 자료행렬(data matrix)을 나타낸다. 또한 은 길이가 인 영벡터(zero vector)이며 은 크기가 인 단위행렬(identity matrix)이다.

문장표절률: 0%

회귀계수벡터의 추정량 은 보통 최소제곱추정법(least squares estimation : LSE)을 이용하여 계산하게 된다.

문장표절률: 23%

이외에도 오차의 정규성(normality)을 가정한 최대가능도추정법(maximum likelihood estimation : MLE)을 사용할 수도 있지만 다중회귀분석의 경우는 이 2가지 방법에 의해 계산된 추정량의 형태가 결과적으로는 동일하다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

의 형태를 만족한다는 가정 하에 최대가능도추정법(maximum likelihood estimation : MLE)을 통해서 추정량 을 구하게 된다

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

의 형태를 만족한다는 가정 하에 최대가능도추정법 maximum likelihood estimation : MLE)을 통해서 추정량 3을 구하게

문장표절률: 0%

최소제곱추정법을 적용하는 계산과정은 아래의 식과 같으며 이러한 과정을 통해 얻어지는 식을 정규방정식(normal equation)이라고 한다.

문장표절률: 28%

만약 위의식에서 행렬에 대한 역행렬(inverse matrix)이 존재하는 경우, 다시 말해서 행렬 가 완전계수(full-rank)의 성질을 만족하게 된다면 의 추정량 은 다음과 같은 형태의 유일한 하나의 해로 결정되며, 이 추정량에 대한 기댓값(expected value)과 분산(variance)은 다음과 같이 증명할 수 있다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

을 얻을 수 있다. 만약 행렬 가 완전계수(full-rank)의 성질을 만족하는 경우, 다시 말해서 행렬 의 역행렬(inverse matrix)이 존재하는 경우 의 추정량 은 다음과 같이

문장표절률: 0%

결론적으로 추정량 은 회귀계수벡터 의 비편향추정량(unbiased estimator)이라는 사실을 확인할 수 있다.

문장표절률: 0%

참고로 최대가능도추정법을 사용하여 회귀계수벡터 의 추정량 를 계산하면 다음과 같다.

문장표절률: 0%

---------- If , then probability distribution function(pdf) of vector is

문장표절률: 0%

. And . Therefore, pdf of vector is . Hence, likelihood function of and is... So, if matrix is full-rank, maximum likelihood estimator of and are...

문장표절률: 0%

and *----------* 3.2 능형 회귀분석 (ridge regression)

문장표절률: 0%

능형 회귀분석(ridge regression)은 라소 회귀분석(lasso regression : Least Absolute Shrinkage Selector Operator)과 함께 대표되는 벌점회귀(penalized regression) 방법의 일종이다.

문장표절률: 26%

이 방법은 데이터 분석 시 3.1절에서 소개한다중회귀분석을 적용하고자 할 때 설명변수들 사이에 연관성이 존재하여 발생하게 되는 다중공선성의 문제(multicollinearity problem)가 있다고 여겨지는 경우, 또는 관측치의 개수보다 설명변수의 개수가 더 많아서 추정량이 유일하게 결정되지 않는 경우 등의 상황에서 자주 적용되는 방법이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

때 다중공선성의 문제가 있다고 여겨지거나 관측치의 개수보다 설명변수의 개수가 더 많아서 추정량이 유일하게 결정되지 않는 경우에 자주 사용하는 방법이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

때 다중공선성의 문제가 있다고 여겨지거나 관측치의 개수보다 설명변수의 개수가 더 많아서 추정량이 유일하게 결정되지 않는 경우에 자주 사용하는 방법이다. 이 방법의

문장표절률: 0%

이 방법의 핵심은 회귀계수의 추정량을 구하기 위하여 어느 정도의 편의(bias)를 허용하여 비편향추정량의 이점을 약간 포기하는 대신 분산(variance)을 큰폭으로 줄일 수 있도록 적절하게 규제(regulation)하여 좀 더 신뢰성이 높은 추정량을 얻게 된다는 것에 있다.

문장표절률: 26%

추정량은 다음과 같이 최소제곱추정법과 비슷한 방법을 통하여 계산하게 된다. 다만 일반적인 다중회귀분석과는 다르게 양의 실수가 곱해진 이차형식 형태의 벌점함수(penalty function) 를 추가로 적용하여 추정량을 구한다는 것이 능형 회귀분석의 큰 특징이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

된다. 추정량은 다음과 같은 방식으로 양의 실수가 곱해진 이차형식 형태의 벌점함수(penalty function) ‖ ‖ 를 적용하여 구

문장표절률: 0%

여기에서는 능형모수(ridge parameter)를 의미하며 능형 회귀분석에 의해 계산되는 추정량의 편의와 분산의 비율을 적절하게 잡아주는 중요한 역할을 한다.

문장표절률: 32%

만약의 값이 에 가깝게 되면 추정량의 편의는 에 가깝게 되고 분산은 커지게 된다. 하지만 반대로 의 값이 증가하게 되면 편의는 커지게 되지만 분산은 줄어들게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

ridge parameter)로써 만약 이 값이 에 가깝게 되면 추정량의 편의는 에 가깝게 되지만 분산은 커지게 된다. 하지만

문장표절률: 20%

그러므로 능형 회귀분석을 통해 추정량을 계산하기 위해서는 사전에 적절한 모수 의 값을 정해주어야 하며 보통 교차타당성(cross validation : CV)의 방법을 이용하여 test MSE(mean squared error)의 추정값이 가장 작게 나오는 상황을 만들어주는 의 값을 가장 바람직한 조건으로 판단하고 선택하게 된다. 이 방법을 통하여 얻어지는 의 추정량은 다음과 같다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

보통 CV(cross validation) 를 이용하여 test MSE(mean squared error)의 추정값을 가장 작게 만들어주는 A값을 최적의 조건으로

[Copykiller] 커널능형회귀분석에서 앙상블기법을 이용한 효율성 연구 = A study on kernel ridge regression using ensemble method

저자 : 韓善雨
발행 : 韓國外國語大學校 大學院 : 통계학과 2016. 2

일반적으로 CV(cross validation)를이용해 test MSE(mean squared error)의

문장표절률: 0%

이러한 추정량의 기댓값은 회귀계수벡터 와는 다르다. 따라서 에 대한 비편향성(unbiased)을 만족하지는 않는다. 하지만 위의 식에서 행렬은 에 의해서 반드시 역행렬을 가지게 된다.

문장표절률: 0%

따라서 추정량 은 의 값에 따라 유일하게 하나의 값으로 결정이 된다. 참고로 라소 회귀분석에서 사용하는 벌점함수의 형태는 이며 의 값에 따라 추정된 회귀계수의 값이 이 되는 경우도 존재한다.

문장표절률: 0%

이러한 특성 때문에 라소 회귀분석은 변수선택(variables election)의 기능을 포함하고 있으며 특정 설명변수에 대한 회귀계수를 추정하기 위해 나머지 설명변수들에 대한 회귀계수들이 주어졌다고 가정하고 회귀계수를 추정해나가는 soft-threshold ing의 방법을 통해 추정량을 계산한다.

문장표절률: 0%

이에 대해서는 Kim, J. (2018)을 참조하기 바란다. 그리고 능형 회귀분석과 라소 회귀분석을 절충하여 혼합한 방법(hybrid method)인 엘라스틱 넷 회귀분석(elastic-net regression)도 존재하며이는 극단적으로 변수의 개수가 관측치의 수보다 많거나 다중공선성의 문제가 강하게 존재한다고 할 때 선호된다.

문장표절률: 0%

벌점회귀(penalized regression)에 대한 자세한 내용은 Friedman et al. (2007)과 Hastie et al. (2011)을 참조하기 바라며 각 방법에 대한 핵심을 정리하면 다음과 같다.

문장표절률: 0%

----------- Ridge regression : Lasso regression : Elastic-net regression : *-----------* 3.3 커널 능형 회귀분석 (kernel ridge regression)

문장표절률: 0%

3.2절에서 소개한 능형 회귀분석(ridge regression) 방법은 반응변수와 설명변수들 사이에선형(linear relationship)관계가 있다는 가정을 기본으로 분석을 진행한다.

문장표절률: 0%

하지만 우리가 실질적으로 수집하고 분석하게 되는데이터들은 대부분 처리가 원활하도록 정리되어 있지 않기 때문에 사전에 연구자가 분석의 목적에 맞춰서 정리해주는 절차가 필요하다.

문장표절률: 0%

거의 비선형(nonlinear relationship)구조를 보이는 것이 일반적이기 때문에 적절한 변환함수(transformation function)에 대한 고려가 필요하며 때에 따라서는 설명변수들 사이에 있는 상호작용(interaction effect) 등을 생각해야 하는 경우도 빈번하게 발생한다.

문장표절률: 40%

이러한 문제점들이 있는 경우에 적용할 수 있는 방법이 바로 커널트릭 기법(kernel-trick method)이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

경우에 적용할 수 있는 방법이 커널트릭 기법(kernel-trick method)이다. 이 기법은 복잡 한 형태를

문장표절률: 21%

이 방법을 적용함으로써 복잡한 구조를 보이는 비선형 데이터에 대하여 이 데이터의 특성에 맞는 적절한 사상함수(mapping function) 를 적용해 차원 설명변수 공간에 존재하는 데이터를 특성에 맞게 변형시킬 수 있고, 그 결과는 고차원의 힐버트 공간(Hilbert space) 또는 특성공간(featurespace)에 놓이게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

형태를 나타내는 비선형 데이터에 대해 적절한 사상함수 mapping function) 를 적용해 차원의 설명변수 공간에 있는 데이터를 변형하여 고차원의 힐버트

문장표절률: **18%**

이러한 과정을 통해서 데이터를 변형시키게 되면 미리 변환함수를 고려하지 않더라도 그 데이터의 특성에 맞는 변환함수를 적용한 것과 같은 결과를 적절하게 얻을 수 있으며 변형된 데이터에 대해 선형모형(linear model)을 적합하여 분석을 진행하게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

변환함수를 적용한 것과 같은 효과를 얻을 수 있으며 변형된 데이터에 대해 선형 14 〈그림 1

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

변환함수를 적용한 것과 같은 효과를 얻을 수 있으며 변형된 데이터에 대해 선형 14 모형을 적합하여 분석을

문장표절률: **0%**

커널트릭 기법의 핵심을 간략하게 표현한다면 아래에 있는 〈그림 2〉와 같이 나타낼 수 있다.

문장표절률: **0%**

즉, 원데이터(original data)에서는 비선형 관계를 보이기 때문에 그 자체를 그대로 두고 분석한다고 했을 때는 적절한 모형을 적합하기 어려운 문제 라고 할지라도 이를 적절하게 변형하여 특성공간으로 이동시키게 되면 그 관계가 선형으로 바뀌기 때문에 모형을 적합하기 쉬운 문제로 바뀌게 된다는 것이 커널트릭 기법의 핵심 이라고 할 수 있겠다.

문장표절률: **0%**

〈그림 2 : Kernel-trick method〉 지금부터는 이러한 커널트릭 기법을 어떻게 능형 회귀분석방법에 적용할 수 있는지 설명하며 더 자세한 내용은 Huh, M.

문장표절률: **0%**

(2015), Lee et al. (2016), Han, S. (2016), 그리고 Hwang, S. (2017)을 참조하기 바란다. 개의 관측치와 차원의 설명변수 공간을 가진 훈련자료(training data) 가 있다고 하고 이 훈련자료에 있는 개의 관측치가 이라고 가정해보자.

문장표절률: **0%**

이에 대하여 사상함수를 이용하여 다음과 같은 방법을 통해 훈련자료에 대한 개의 설명변수 데이터들을 적절하게 변환할 수 있다.

문장표절률: **35%**

사상함수를 이용하여 변환된 설명변수 데이터 은 고차원의 특성공간(featurespace with high dimension)에 놓이게 되며, 이를 이용하여 다음과 같은 회귀모형을 적합할 수 있다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

Hilbert space)이라고 불 리는 고차원의 특성공간(featurespace with high dimension)에 놓이게 된다. 이를 이용하여 다음과 같은

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

Hilbert space)이라고 불 리는 고차원의 특성공간(featurespace with high dimension)에 놓이게 된다. 이를 이용하여 다음과 같은

문장표절률: **0%**

위와 같은 모형을 적합하는 과정을 통하여 길이가 인 회귀계수벡터를 설정할 수 있다.

문장표절률: **0%**

이러한 커널트릭을 통한 공간변환은 실제적으로 커널함수에 의한 계산을 통해서 이루어진다.

문장표절률: **0%**

이러한 과정을 거침으로써의 선형결합(linear combination) 에 대한 의사영(projection)은 다음과 같이 계산된다.

문장표절률: 0%

여기에서는 행렬의 번째 원소이다. 이러한 계산과정을 통해서 얻어지게 되는 행렬 와 회귀계수벡터를 이용해서 반응변수를 설명하게 된다.

문장표절률: 0%

이에 대한 자세한 내용은 Schölkopf and Smola (2002)를 참조하기 바란다. 행렬 를 계산하기 위해 사용하는 커널은 다음과 같은 머서의 정리(Mercer's theorem)를 만족하여야 하며 이는 Minh et al. (2006)와 Nguyen, V. (2015)에 자세하게 설명되어 있다.

문장표절률: 66%

-- Mercer's theorem A symmetric function can be expressed as an inner product for some if and only if is positive semi-definite, is equal to, for all or, equivalently is positive semi-definite matrix

[Copykiller] Mahalanobis kernel-based support vector data description for detection of large shifts in mean vector

저자 : ["Vu Nguyen"]
발행 : 01/2015

s theorem is stated below. Mercer's theorem A symmetric function can be expressed as an inner product (,) = ⟨ ( ), ( )⟩ (2. 12) for

[Copykiller] [IEEE 2014 International Conference on Mechatronics and Control (ICMC) - Jinzhou, China (2014.7.3-2014.7.5)] 2014 International Conference on Mechatronics and Control (ICMC) - A classifier fusion method based on classifier accuracy

저자 : Li, Wenxing; Hou, Jian; Yin, Lizhi
발행 : 2014

symmetric function can be indicated as an inner product for some if and only if is positive semi-definite. So every kernel matrix is

문장표절률: 0%

for any collection *--* 다시 말해서 커널함수가 내적(inner product)형태의 연속형 함수(continuous function)일 때, 커널함수의 값으로 만든 행렬 가 대칭행렬(symmetric matrix)이면서 준양정치행렬(positive semi-definite matrix)이라면 를 만족하는가 존재한다는 것이 머서의 정리이다.

문장표절률: 0%

이 머서의 정리를 만족하는 커널의 형태는 〈표 2〉에 제시된 바와 같이 다양하게 존재하며이에 대한 자세한 사항은 Karatzoglou et al. (2006)과 Souza, C.R. (2010) 등을 통해서 확인할 수 있다.

문장표절률: 33%

Linear kernel Polynomial kernel Gaussian (Radial Basis) kernel Laplace kernel ANOVA kernel Sigmoid kernel Rational Quadratic kernel Multiquadric kernel Inverse Multiquadric kernel Bessel kernel Cauchy kernel Generalized T-Student kernel

[wiki.math.uwaterloo.ca] stat841f10 - statwiki

발행 : wiki.math.uwaterloo.ca

Kernel ANOVA Kernel, Hyperbolic Tangent (Sigmoid) Kernel Rational Quadratic Kernel Multiquadric Kernel Inverse Multiquadric Kernel Circular Kernel Spherical Kernel, Wave

문장표절률: 0%

Power kernel conditionally positive definite Log kernel conditionally positive definite Triangular kernel positive definite in Circular kernel positive definite in Spherical kernel positive definite in 〈표 2 : 머서의 정리를 만족하는 다양한 커널의 형태〉

문장표절률: 31%

본 연구에서는 머서의 정리를 만족하는 커널들 중 다항커널(Polynomial kernel)과 가우시안 커널(Gaussian kernel)을 적용한다. 이 2가지 커널은 다음과 같다. * Polynomial kernel : * Gaussian kernel :

[hoon427.tistory.com] 서포트 벡터 머신

발행 : hoon427.tistory.com

부른다. 커널 함수의 대표적인 예에는 다항커널(Polynomial Kernel)과 가우시안 커널(Gaussian Kernel,레이디얼 베이스 함수 커널Radial

[Copykiller] GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : ...

저자 : 박영욱 GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : the case of domestic 5 stars hotel reviews on Trip Advisor web site

주성분 분석에서 커널 함수 설정은 다항커널(polynomial kernel)과 가우시안 커널(gaussian kernel)중에서 가우시안 커널을 사용하였다.t

문장표절률: 0%

(단,) 특히 가우시안 커널은 다른 커널들과 비교했을 때 분석하고자 하는데이터에 대한 사전정보(prior information)가 알려져 있지 않은 경우에도 유연하게(flexible) 적용이 가능하다는 강력한 특성을 가지고 있다.

문장표절률: 26%

그 이유는 가우시안 커널의 경우 다음과 같이 테일러 급수 전개(Taylor series expansion)를 통해 무한 급수의 합으로 나타낼 수 있고 이는 결국 차원이 무한대인 벡터 형태의 변환된 설명변수와 의 내적으로 표현이 가능하다는 것을 의미한다.

[www.jksmer.or.kr] Considerations in the Estimation of Rock Mass Strength by ...

저자 : Dohyun Park

추정한다. 이러한 다항식 형태의 추정은 테일러 급수 전개(Taylor series expansion)를 통해 어떤 함수를 유한차수(n차

[jksmer.or.kr] Considerations in the Estimation of Rock Mass Strength by ...

추정한다. 이러한 다항식 형태의 추정은 테일러 급수 전개(Taylor series expansion)를 통해 어떤 함수를 유한차수(n 차

문장표절률: 0%

즉, 가우시안 커널을 적용하게 된다면 분석하고자 하는 데이터를 적절하게 변형하여 무한차원(infinite space)의 특성공간(featurespace)으로 이동시키게 된다.

문장표절률: 0%

본 연구에서는 다항커널을 적용하는 경우에는 커널로 인해 만들어지는 경계(boundary)의 유연성(flexibility)을 결정해주는 정도모수(degree parameter) 를 으로 고정하였는데 그 이유는 설명변수의 개수가 클수록 복잡도가 더 커지는 만큼 이를 보완하기 위해 의 크기를 작게 해야 하는데, 이라고 고정하는 것도 이에 대한 충분한 해결책이 된다고 판단했기 때문이다.

문장표절률: 0%

같은 논리를 바탕으로 규모모수(scale parameter) 는 로 설정하였으며 차감모수(offset parameter) 는 로 고정하였다.

문장표절률: 0%

그리고 가우시안 커널을 적용하는 경우에는 설명변수의 개수가 클수록 boundary를 간단한 형태로 설정하기 위해 커널로 인해 만들어지는 경계의 유연성을 결정해주는 조절모수(tuning parameter) 를 로 설정하였다.

문장표절률: 0%

여기에서는 커널함수를 통하여 변환하기 전 상태의 원데이터가 놓여 있는 설명변수 공간의 차원을 뜻한다.

문장표절률: 0%

이와 같은 커널 변환을 이용하여 아래와 같은 형태의 회귀모형을 얻을수 있다.

문장표절률: 0%

그리고 변환을 통해서 얻어진 행렬에 대한 역행렬이 항상 존재하지 않는다는 점을 보완하기 위해 능형 회귀분석 형태의 벌점함수를 적용하여 회귀계수벡터를 추정한다. 이를 통해 추정된 회귀계수벡터 는 다음과 같이 계산된다.

문장표절률: 0%

여기에서 최적의 능형모수 의 값은 겹 교차타당성(-fold cross validation)을 통하여 선정한다.

문장표절률: 0%

본 연구에서는 로 설정하였고 최적의 값은 RMSE(Root Mean Square Error)의 평균값이 최소인 경우를 최적의 상황으로 판단하고 선택하였다.

문장표절률: 23%

이러한 방법을 통하여 훈련자료(training data) 를 통한 검증자료(test data) 에 대한 평가를 실시하기 위해서는 커널함수를 사용하여 변환된 설명변수 데이터의 선형결합인 에 대한 의사영(projection)을 계산한 뒤 이를 바탕으로 검증자료 평가에 사용할 행렬 를 계산해야 하며, 이에 대한 계산과정은 다음과 같다. 단, 은 훈련자료의 관측치 개수, 는 검증자료의 관측치 개수를 뜻한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

훈련자료(training data) 를 이용하여 검증자료(test data) 에 대한 평가를 실시하고자 한다면 커널함수를 사용해서 … 의 선형결합인 … 에 대한 의사영(projection)을 계산한 다음 이를 바탕으로 검증자료 평가를

문장표절률: **50%**

그리고 본 연구에서는 다음과 같은 형태의 다항커널(Polynomial kernel)과 가우시안 커널(Gaussian kernel)을 적용한다.

[hoon427.tistory.com] 서포트 벡터 머신

발행 : hoon427.tistory.com

부른다. 커널 함수의 대표적인 예에는 다항커널(Polynomial Kernel)과 가우시안 커널(Gaussian Kernel,레이디얼 베이스 함수 커널Radial

[Copykiller] GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : ...

저자 : 박영욱 GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : the case of domestic 5 stars hotel reviews on Trip Advisor web site

주성분 분석에서 커널 함수 설정은 다항커널(polynomial kernel)과 가우시안 커널(gaussian kernel)중에서 가우시안 커널을 사용하였다.t

문장표절률: **0%**

* Polynomial kernel : * Gaussian kernel : (단,) 여기에서는 행렬의 번째 원소이며 는 검증자료 안에 있는 번째 관측치에 대한 설명변수 데이터를 의미한다.

문장표절률: **0%**

이러한 과정을 통하여 얻을수 있는 행렬 와 훈련자료를 통해서 계산된 추정량을 이용하여 검증자료에 대한 최종적인 평가를 진행 할 수 있다.

문장표절률: **0%**

다시 말해서 행렬 와 추정량을 이용하여 다음과 같은 형태의 훈련자료를 통하여 계산된 검증자료에 대한 반응변수의 추정량 를 구할 수 있다.

문장표절률: **0%**

이 추정량을 바탕으로 다음과 같은 RMSE 값을 계산하여 이 값이 작을수록 구축한 모형의 예측력이 높다고 판단하게 된다. 3.4 커널 능형 중도절단 회귀분석 (kernel ridge censored regression)

문장표절률: **0%**

커널 능형 중도절단 회귀분석(kernel ridge censored regression)은 3.3절에서 설명한 커널 능형 회귀분석에서 데이터 내의 반응변수인 를 2.3절에서 소개한 인조변수(synthetic response) 로 바꿔서 표현해 주면 된다. 그리고 그 형태는 다음과 같이 표현할 수 있다.

문장표절률: **0%**

, , , , 다시 말해서 분석하고자 하는 생존자료(survival data)에 포함되어 있는 관측된 생존시간 와 중도절단의 여부를 나타내는 지시변수를 사용하여 실제 환자에게서 관심사건이 일어날 때까지 걸리는 시간 를 대체하는 인조변수를 계산한 뒤 이를 커널 능형 회귀분석에 적용하면 되는 것이다.

문장표절률: **39%**

본 연구에서는 함수를 절단변수가 설명변수에 의존하지 않는다는 가정을 바탕으로 카플란-마이어 추정량(Kaplan-Meier estimator)을 사용하여 추정한다.

[boxnwhis.kr] 유저 생존 곡선 그리기 :: -[1]- Box and Whisker

발행 : boxnwhis.kr

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다. 이론적인 생존 곡선과는 다르게

[www.boxnwhis.kr] 유저 생존 곡선 그리기 :: -[1]- Box and Whisker

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다.

문장표절률: **60%**

이외의 다른 사항들은 커널 능형 회귀분석의 내용과 동일하기 때문에 이에 대한 자세한 설명은 생략하도록 한다.

[clomag.co.kr] [신광섭의 데이터바로보기]데이터과학자, 네 정체가 뭐니 - CLO

발행 : clomag.co.kr

사업에 대한 지식' 역시 마찬가지다. 때문에 이에 대한 자세한 설명은 생략하도록 한다.

[patents.google.com] KR102175041B1 - 가상 시리얼 포트를 이용한 캡티브 포탈 주문 결제 출력 시....

발행 : 2019-07-19, 2020-11-05

결제하는 방법은 통상의 방법과 같으므로, 이에 대한 자세한 설명은 생략하도록 한다. 한편, 사업장이 제공하는 와이파이를 이용한

문장표절률: 0%

다만, 커널 능형 중도절단 회귀분석에서는 인조변수를 추가로 사용하기 때문에 구축된 모형을 평가시 예측력의 평가기준을 인조변수로 하는 경우와 실제 반응변수로 하는 경우로 나누어서 살펴봐야 한다.

문장표절률: 0%

다시 말해서 평가기준을 인조변수로 하는 경우에는 RMSE를 로 계산하고, 평가기준을 실제 반응변수로 하는 경우에는 RMSE를 로 계산하면 된다.

문장표절률: 0%

다만 실제 생존자료에서는 모든 환자의 실제 생존시간, 즉 관심사건이 일어날 때까지 걸리는 시간이 기록되어 있지 않다는 점을 감안하여 본 연구에서는 임의로 데이터를 생성할 수 있는 모의실험에서는 위에서 설명한 2가지 기준을 모두 살펴보고 실제 데이터를 바탕으로 실시하는 실증분석에서는 평가기준을 인조변수로 하는 경우에 한하여 방법론의 성능을 평가할 것이다.

문장표절률: 0%

4 앙상블 기법 (ensemble method) 본 장에서는 부트스트랩(bootstrap) 기법을 응용하여 추정량의 변동성을 큰폭으로 줄일수 있는 방법인 앙상블 기법(ensemble method)을 소개한다.

문장표절률: 18%

앙상블 기법은 기계학습(machine learning)의 한 종류인 나무모형 기법(tree-model method)에서 유래한 방법으로 여러 개의 예측기(predictor) 또는 분류기(classifier)를 생성한다음 이를 결합하여 보다 더 정확한 결과를 도출하는 기법이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

기계학습(statistical machine learning)의 나무모형 기법(tree-model method)에서 나온 분석방법이다. 본 연구에서는 배깅

문장표절률: 57%

본 연구에서는 다양한 앙상블 기법 중 배깅(bagging)과 랜덤포레스트(random forests)를 사용한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

에서 나온 분석방법이다-. 본 연구에서는 배깅 bagging)과 랜덤포레스트 random forests)를 사용한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

에서 나온 분석방법이다. 본 연구에서는 배깅(bagging)과 랜덤포레스트 random forests)를 사용한다. 이 2가지 방법의 공통적인

문장표절률: 0%

이 2가지 방법론이 공유하는 핵심은 서로 독립적으로 추출된 부트스트랩 표본들을 이용하여 얻은 결과들을 평균한 값을 추정량으로 사용하여 분산을 큰폭으로 줄여서 예측에 대한 안정성을 도모한다는 것에 있다.

문장표절률: 24%

이러한 방법을 적용한다면 보다 더 정확하고 신뢰성 있는 예측을 할 수 있다. 배깅과 랜덤포레스트 외에도 AdaBoost(Adaptive Boosting), GBM(Gradient Boosting Machine), XGBoost, LightGBM 등의 다양한 부스팅 알고리즘(Boosting Algorithm)도 앙상블 기법에 속하며 순차적으로(sequentially) 여러 개의 약한 학습기(weak learner)들을 결합하여 모형의 성능을 높이는 것이 핵심이다.

[sevillabk.github.io] [사이킷런] 부스팅 알고리즘(AdaBoost & GBM) - SevillaBK's Blog

알고리즘(Boosting Algorithm) 부스팅 알고리즘은 여러 개의 약한 학습기(weak learner)를 순차적으로 학습-예측하면서 잘못 ..... 아래와 같은 알고리즘들이 있습니다. AdaBoost GBM(Gradient Boosting Machine) XGBoost LightGBM Cat Boost AdaBoost AdaBoost는

[Copykiller] 기계학습의 산업안전 적용에 관한 연구 = A Study on Application of Machine Learning for Occupational Safety

저자 : 최승주
발행 : 2021

3.3 XGBoost 부스팅 기법은 여러 개의 약한 학습기(Weak learner)를 순차적으로 학습시키고 이를 결합하여

문장표절률: 0%

이에 대해서는 Freund et al. (1999), Lee, J. (2020), Han, S. (2016), 그리고 Hwang, S. (2017)을 참고하기 바란다.

문장표절률: 0%

간단한 예시를 보도록 하자. 만약 임의의 개의 표본이 서로 독립이고 분산이 모두 으로 같다고 가정한다면 이들의 표본평균(sample mean) 의 분산은으로 크게 줄어들게 된다.

문장표절률: 0%

즉, 많은 수의 표본들을 추출후 이를 평균하여 얻은 추정량은 그러한 과정을 거치지 않고 일반적인 방법을 통해서 얻은 추정량과 비교했을 때 분산의 크기가 매우 작기 때문에 변동성이 크게 줄어들게 된다.

문장표절률: 0%

그러므로 평균화를 통해서 얻은 추정량은 신뢰성 측면에서 일반적인 추정량보다 더 좋은 성능을 보이게 된다.

문장표절률: 38%

그리고 수식의 형태를 통해 표본의 개수 과 분산은 서로 반비례 한다는 사실을 쉽게 파악할 수 있다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

바람직한 성능을 보이게 된다. 그리고 표본의 개수 과 분산은 서로 반비례 한다. 그러므로 신뢰성 측면에서 좀

문장표절률: 0%

그러므로 일반적인 추정량과 비교했을 때 신뢰성 측면에서 더 바람직한 추정량을 구하고 싶다면 많은 개수의 표본들을 추출해서 각각에 대한 추정량을 구하고 이들을 평균하는 과정을 통해 분산의 값을 큰폭으로 줄이면 되는 것이다.

문장표절률: 0%

하지만 그렇다고 표본의 개수를 무작정 무한대에 가까울 정도로 크게 늘린다는 것은 현실적으로 불가능한 일이다.

문장표절률: 0%

그러므로 원래의 데이터에서 충분한 개수의 부트스트랩 표본을 뽑아서 분석하는 배깅이나 랜덤포레스트와 같은 방법은 이러한 한계를 어느 정도 극복하게 해주는 분석방법 이라고 말할 수 있다.

문장표절률: 16%

4.1 배깅 (bagging) 배깅(bagging : bootstrap aggregation)은 하나의 훈련자료(training data)에 대하여 반복적인 복원추출(sampling with replacement)을 실시함으로써 훈련자료와 비교하여 관측치의 개수가 동일한 여러 개의 부트스트랩(bootstrap) 표본들을 추출하여 분석하는 앙상블 기법의 한 방법이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

하나의 훈련자료 training data)에서 반복적인 복원추출 sampling with replacement)을 통하여 훈련자료와 비교했을 때 관측치의

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

하나의 훈련자료(training data)에서 반복적인 복원추출(sampling with replacement)을 통하여 훈련자료와 비교했을 때 관측치의

문장표절률: 0%

추출된 부트스트랩 표본에 대하여 모두 같은 유형의 알고리즘(algorithm) 기반의 예측기(predictor) 또는 분류기(classifier)를 사용하여 분석하며, 모형을 적합할 시 발생할 수 있는 과대적합(overfitting)의 문제를 보완할 수 있다는 장점이 있다.

문장표절률: 0%

여기에서 과대적합이란 기계학습(machine learning)에서 훈련자료를 과도하게 학습하게 되어 훈련자료에 대해서는 오차가 감소하지만 전체 자료에 대해서는 오히려 오차가 증가하는 현상을 의미한다.

문장표절률: 0%

반대로 훈련자료에 대한 학습이 부족하여 적절하지 않은 모형이 적합되는 현상도 발생할수 있는데 이를 과소적합(underfitting)이라고 부른다.

문장표절률: **54%**

이에 대한 자세한 사항은 Hastie et al. (2011) 및 James et al. (2014)를 참조하기 바란다. 〈그림 3〉은 부트스트랩 기법에 관하여 표현한 것이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

기법과 부트스트랩에 대한 자세한 내용은 Hastie et al. (2011) 및 James et al. (2014)를 참조하기 바란다. 그림 2는 부트스트랩 기법에 관하여

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

자세한 내용은 Hastie et al. (2011) 및 James et al. (2014)를 참조하기 바란다. 그림 2는 부트스트랩 기법에 관하여

문장표절률: **0%**

이 그림을 통해서 알 수 있듯이 부트스트랩의 핵심은 훈련자료에서 다수의 부트스트랩 표본을 추출한 뒤 각각의 표본을 바탕으로한 추정량을 모두 구한다음 이를 적절하게 조합하여 단순하게 훈련자료를 한번 사용해서 구한 추정량과 비교했을 때보다 더 성능이 좋은 추정량을 계산한다는 것이다.

문장표절률: **0%**

그리고 〈그림 4〉는 과대적합과 과소적합에 대한 개념을 간단하게 표현한 것이다. 이 그림을 통해 모형을 적합 시 과대적합과 과소적합을 피하도록 알고리즘을 구현하는 것이 중요함을 확인할 수 있다.

문장표절률: **0%**

〈그림 3 : Bootstrap〉 〈그림 4 : 과대적합과 과소적합〉 예를 들어 다음과 같이번의 부트스트랩을 통해서 얻은 각각의 표본에 대하여 계산한 개의 추정량이 있다고 가정하자. 배깅 추정량은 이 개의 추정량들을 평균하여 구할 수 있다.

문장표절률: **0%**

이러한 과정을 통하여 상대적으로 분산이 작은 다음과 같은 형태의 배깅 추정량을 만들 수 있다.

문장표절률: **36%**

4.2 랜덤포레스트 (random forests) 랜덤포레스트(random forests)는 배깅의 문제점을 보완하기 위해 제안된 앙상블 기법의 한 방법으로 추정량을 구하는 전체적인 과정은 배깅의 원리와 비슷하다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

수 있게 된다. 27 3.2 랜덤포레스트 (random forests) 랜덤포레스트(random forests)는 배깅 기법의 문제점을 보완하기 위해 제안된

문장표절률: **25%**

이 방법도 4.1절에서 소개한 배깅의 경우와 비슷하게 부트스트랩 기법을 이용하여 여러 개의 추정량들을 얻은 뒤 이들에 대한 평균을 계산하여 최종적인 추정량을 얻게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

이 기법에서도 베깅 기법과 동일하게 부트스트랩 기법을 이용하여 여러 개의 추정량들을 만들고 이들을 평균내서 최종적인 추정량을

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

이 기법에서도 배깅 기법과 동일하게 부트스트랩 기법을 이용하여 여러 개의 추정량들을 만들고 이들을

문장표절률: **25%**

하지만 각각의 표본들에 대해 모든 설명변수를 다 사용하지 않고 이들중 일부만을 선택해서 사용한다는 특징이 모든 설명변수를 사용하는 배깅과 비교했을 때 예측력과 신뢰성을 높이는 큰 차이를 주게 되는 것이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

각각의 표본들에 대해 모든 설명변수들을 다 사용하지 않고 이들중 일부만을 선택해서 사용한다는 점이 예측력을 높이는 큰 차이를

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

각각의 표본들에 대해 모든 설명변수들을 다 사용하지 않고 이들중 일부만을 선택해서 사용한다는 점이 예측력을 높이는 큰 차이를

문장표절률: **0%**

부트스트랩 표본에 대해 모든 설명변수를 다 포함시켜 추정량을 구하는 배깅 기법을 사용하게 된다면 각각의 부트스트랩 표본들에 의하여 만들어진 하위모형들(submodels) 사이에 서로 연관성(correlated)이 존재할 수 있기 때문에 추정 시 정확도가 떨어질 수 있다.

문장표절률: **0%**

이러한 단점을 보완하기 위해 모든 설명변수 중 일부만을 선택하여 각각 부트스트랩 표본에 적용함으로써 그 연관성을 크게 낮출 수 있다.

문장표절률: **0%**

물론 이러한 과정을 거치게 되면 불가피하게 편의(bias)가 발생하게 된다. 하지만 그만큼 연관성을 크게 줄이게 되므로 결과적으로는 분산을 좀 더 큰폭으로 줄일수 있기 때문에 이를 통하여 편의에 대한 효과를 상쇄시켜줄일 수 있게 된다. 이러한 방법을 통하여 좀 더 정확한 예측을 할 수 있다.

문장표절률: **0%**

예를 한번 살펴보도록 하자. 만약 설명변수의 개수가 인 훈련자료가 있다고 하고 이 훈련자료를 이용하여 번의 부트스트랩을 통해서 개의 표본들을 추출했다고 가정해보자.

문장표절률: **43%**

여기에서 랜덤포레스트의 경우는 배깅의 경우와는 다르게 각각의 표본을 추출할 때마다 개의 모든 설명변수들 중 개만을 선택해서 부트스트랩 표본에 포함시켜야 한다는 점에 주의해야 한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

개의 표본들을 추출한 다. 여기에서 각각의 표본을 추출할 때마다 개의 모든 설명변수들 중 개만을 선택해서 포함시켜야 한다는 점에 주의할 필요가

문장표절률: **26%**

일반적으로의 경우는 또는 라고 설정하게 되는데 본 연구에서는 를 적용하였으며 의 값이 자연수의 형태로 나타나지 않는 경우에는 소수점 이하 반올림을 사용하여 부트스트랩 표본에 들어갈 설명변수의 개수를 결정하였다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

정하게 된다. 본 연구에서는 ≈ 를 적용하였으며 의 값이 자연수의 형태로 나타나지 않을 경우에는 소수점 이하 반올림을

문장표절률: **40%**

그리고 각각의 부트스트랩 표본에 들어갈 개의 설명변수들은 부트스트랩 표본을 추출할 때마다 무작위로 다르게 선택해야 한다는 사실에도 주의를 기울여야 할 필요가 있다. 이러한 과정을 통해서 다음과 같은 개의 추정량을 계산하게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

개의 설명변수들은 부트스트랩 표본을 추출할 때마다 다르게 선택해야 한다는 점도 유념해야 할 것이다. 이러한 과정을 통해서 다음과 같은 개의 추정량들을 얻을 수 있다. …

문장표절률: **0%**

랜덤포레스트 추정량은 이 개의 추정량들을 평균하여 구할 수 있다. 이를 통해 배깅 기법의 경우와 비교했을 때 좀 더 정확도가 높은 우수한 다음과 같은 형태의 랜덤포레스트 추정량을 얻을수 있다.

문장표절률: **0%**

이러한 과정을 통해서 얻은 랜덤포레스트 추정량은 배깅 기법에 의해서 계산된 추정량과 비교했을 때 정확성, 그리고 신뢰성 측면에서 더 바람직한 성능을 나타내게 된다.

문장표절률: **0%**

이어지는 5장에서는 배깅이나 랜덤포레스트 같은 앙상블 기법을 어떻게 커널 능형 중도절단 회귀분석에 적용할 수 있는지 소개한다.

문장표절률: **0%**

5 앙상블 기법을 이용한 커널 능형 중도절단 회귀분석 본 장에서는 4장에서 소개한 앙상블 기법을 어떻게 커널 능형 중도절단 회귀분석에 적용할 수 있는지 설명한다.

문장표절률: **0%**

4장에서도 언급했듯이 앙상블 기법은 다수의 독립적인 부트스트랩 표본을 사용하여 추정량의 분산과 설명변수들 사이의 연관성을 큰폭으로 줄임으로써 추정의 정확도와 신뢰도를 크게 높일수 있는 방법론이다.

문장표절률: **0%**

이를 커널 능형 중도절단 회귀분석에 적용한다면 보다 더 환자의 생존시간, 다시 말해서 관심사건이 일어날 때까지 걸리는 시간을 정확하게 예측하는 모형을 구축할 수 있게 된다.

문장표절률: **18%**

5.1 배깅 기법을 이용한 커널 능형 중도절단 회귀분석관측치의 개수가 이고 설명변수 공간이 차원인 중도절단이 포함된 훈련자료(training data with censoring) 가 있다고 하고 이 자료에 다음과 같은 개의 관측치가 포함되어 있다고 가정해보자.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

수 있는지 소 개한다. 4.1 배깅 기법을 이용한 커널 능형 로지스틱 회귀분류법 관측치의 개수가 이고

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

수 있는지 소 개한다. 4.1 배깅 기법을 이용한 커널 능형 로지스틱 회귀분류법 관측치의 개수가 n

문장표절률: **39%**

본 연구에서는 함수를 절단변수가 설명변수에 의존하지 않는다는 가정을 바탕으로 카플란-마이어 추정량(Kaplan-Meier estimator)을 사용하여 추정한다.

[boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

발행 : boxnwhis.kr

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다. 이론적인 생존 곡선과는 다르게

[www.boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다.

문장표절률: **0%**

, , , 이 자료에 대하여 부트스트랩을 통해 번째 표본 를 추출하였다고 할 때 이 표본 안에다음과 같은 인조변수(synthetic response) 와 설명변수가 있다고 가정해보자.

문장표절률: **0%**

이에 대하여 사상함수를 이용하여 다음과 같이 부트스트랩 훈련자료에 대한 개의 설명변수 데이터들을 변환할 수 있다.

문장표절률: **30%**

사상함수를 이용하여 변환된 설명변수 데이터 은 고차원의 특성공간(featurespace with high dimension)에 위치하게 되며, 이를 이용하여 다음과 같은 회귀모형을 적합할 수 있다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

Hilbert space)이라고 불 리는 고차원의 특성공간(featurespace with high dimension)에 놓이게 된다. 이를 이용하여 다음과

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

Hilbert space)이라고 불 리는 고차원의 특성공간(featurespace with high dimension)에 놓이게 된다. 이를 이용하여 다음과

문장표절률: **0%**

위와 같은 모형을 적합하는 과정을 통해 길이가 인 회귀계수벡터를 설정하게 된다. 이러한 커널트릭을 통한 공간변환은 실제적으로 커널함수에 의한 계산을 통하여 이루어진다.

문장표절률: **0%**

다시 말해서 의 선형결합(linear combination) 에 대한 의사영(projection)은 다음과 같은 계산을 통하여 얻어지게 된다.

문장표절률: **41%**

여기에서는 행렬의 번째 원소이다. 이러한 과정을 통해서 얻을 수 있는 행렬 와 회귀계수벡터를 이용해서 인조변수를 설명하게 된다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

번째 관측치에 대한 설명변수 데이터이다. 이러한 과정을 통해서 얻을 수 있는 행렬 Kh ₩ 그리

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

번째 관측치에 대한 설명변수 데이터이다. 이러한 과정을 통해서 얻을 수 있는 행렬 , 그리 고 부트스트랩 훈련자료 를

문장표절률: **47%**

본 연구에서는 다항커널(Polynomial kernel)과 가우시안 커널(Gaussian kernel)을 적용한다. 이 2가지 커널은 다음과 같다.

[hoon427.tistory.com] 서포트 벡터 머신

발행 : hoon427.tistory.com

부른다. 커널 함수의 대표적인 예에는 다항커널(Polynomial Kernel)과 가우시안 커널(Gaussian Kernel,레이디얼 베이스 함수 커널Radial

[Copykiller] GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : ...

저자 : 박영욱 GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : the case of domestic 5 stars hotel reviews on Trip Advisor web site

주성분 분석에서 커널 함수 설정은 다항커널(polynomial kernel)과 가우시안 커널(gaussian kernel)중에서 가우시안 커널을 사용하였다.t

문장표절률: **0%**

* Polynomial kernel : * Gaussian kernel : (단,) 이와 같은 커널 변환을 통하여 다음과 같은 형태의 회귀모형을 얻을수 있다.

문장표절률: **0%**

그리고 변환을 통해서 계산된 행렬에 대한 역행렬이 항상 존재하지 않는다는 점을 보완하기 위해 능형 회귀분석 형태의 벌점함수를 적용하여 회귀계수벡터를 추정한다.

문장표절률: **32%**

이를 통해 추정된 회귀계수벡터 는 다음과 같이 계산된다. 여기에서 최적의 능형모수 의 값은 훈련자료에서 임의의 개의 부트스트랩 표본을 추출하여 실시하는 out-of-bag(OOB)의 방법을 통하여 결정한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

에서 임의의 개의 부트스트랩 표본들을 추출하여 실시하는 out-of-bag(OOB)의 방법을 통하여 선정한다. 즉, 추출한 표본들 을

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

임의의 i广개의 부트스트랩 표본들을 추출하여 실시하는 out-of-bag(OOB)의 방법을 통하여 선정한디-.

문장표절률: **0%**

즉, 추출한 표본을 새로운 훈련자료로 두고 이 표본을 추출할 때 뽑히지 않은 나머지 관측치를 해당 표본에 대한 평가자료(validation data)로 이용하여 최적의 조건을 찾는 과정을 거치게 된다.

문장표절률: **0%**

여기에서 최적의 조건이란 각 의 값에 대하여 얻어지는 개의 RMSE(root mean squared error)에 대한 평균값이 가장 작게 나오는 경우를 의미한다.

문장표절률: **23%**

이러한 방법을 통하여 훈련자료(training data) 에서 추출한 번째 부트스트랩 훈련자료를 이용하여 검증자료(test data) 에 대한 평가를 실시하기 위해서는 커널함수를 사용해서 변환된 설명변수 데이터의 선형결합인 에 대한 의사영(projection)을 계산한 뒤 이를 이용하여 검증자료 평가에 사용할 행렬 를 계산해야 하며, 이에 대한 계산과정은 다음과 같다. 단, 은 훈련자료의 관측치 개수, 는 검증자료의 관측치 개수이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

만약 훈련자료(training data) 를 이용하여 검증자료(test data) 에 대한 평가를 실시하고자 한다면 커널함수를 사용해서 … 의 선형결합인 … 에 대한 의사영(projection)을 계산한 다음 이를 바탕으로 검증자료 평가를

문장표절률: **50%**

그리고 본 연구에서는 다음과 같은 형태의 다항커널(Polynomial kernel)과 가우시안 커널(Gaussian kernel)을 적용한다.

[hoon427.tistory.com] 서포트 벡터 머신

발행 : hoon427.tistory.com

부른다. 커널 함수의 대표적인 예에는 다항커널(Polynomial Kernel)과 가우시안 커널(Gaussian Kernel,레이디얼 베이스 함수 커널Radial

[Copykiller] GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : ...

저자 : 박영욱 GloVe를 활용한 G-LDA 문장분석과 디리클레 프로세스 군집분석을 통한 호텔 브랜드 이미지 포지셔닝 : 트립어드바이저 웹사이트 국내 5성급 호텔리뷰를 중심으로 = Study about the brand image positioning on PCA, clustering and G-LDA topic modeling using GloVe : the case of domestic 5 stars hotel reviews on Trip Advisor web site

주성분 분석에서 커널 함수 설정은 다항커널(polynomial kernel)과 가우시안 커널(gaussian kernel)중에서 가우시안 커널을 사용하였다.t

문장표절률: **33%**

* Polynomial kernel : * Gaussian kernel : (단,) 여기에서는 행렬의 번째 원소이며 는 검증자료 안에 있는 번째 관측치에 대한 설명변수 데이터이다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

… 는 검증자료 에 포함되어 있는 번째 관측치에 대한 설명변수 데이터이다. 이러한 과정을 통해서 얻을 수

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

는 검증자료 x니□ 포함되어 있는 번째 관측치에 대한 설명변수 데이터이다. 이러한 과정을 통해서 얻을 수

문장표절률: **0%**

이러한 과정을 통해서 얻어지게 되는 행렬 , 그리고 부트스트랩 훈련자료를 통하여 계산된 추정량을 이용해서 검증자료에 대한 최종적인 평가를 진행한다.

문장표절률: 0%

결과적으로 행렬 와 추정량을 이용하여 다음과 같은 형태의 부트스트랩 훈련자료를 이용하여 계산된 검증자료에 대한 인조변수의 추정량을 구할 수 있다.

문장표절률: 0%

이러한 과정을 개의 부트스트랩 훈련자료에 대해 각각 실시하여 총 개의 추정량 를 얻은 뒤 이를 평균하여 검증자료에 대한 배깅 기법을 이용한 인조변수의 추정량을 구할 수 있으며 그 형태는 다음과 같다.

문장표절률: 0%

커널 능형 중도절단 회귀분석에서는 인조변수를 모형 구축시 사용하기 때문에 구축된 모형을 평가시 예측력의 평가기준을 인조변수로 하는 경우와 실제 반응변수로 하는 경우로 나누어서 살펴봐야 한다.

문장표절률: 0%

이에 따라 본 연구에서는 평가기준을 인조변수로 하는 경우에는 RMSE를 로 계산하고, 평가기준을 실제 반응변수로 하는 경우에는 RMSE를 로 계산하여 성능을 평가한다.

문장표절률: 35%

5.2 랜덤포레스트 기법을 이용한 커널 능형 중도절단 회귀분석 배깅 기법을 적용하는 경우와 동일하게 관측치의 개수가 이고 설명변수 공간이 차원인 중도절단이 포함된 훈련자료가 있다고 하고 이 자료에 다음과 같은 개의 관측치가 포함되어 있다고 가정해보자.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

이용한 커널 능형 로지스틱 회귀분류법 관측치의 개수가 이고 설명변수 공간이 차원인 훈련자료 (training data) 가 있다고

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

계소]:한11牛. 4.2 랜덤포레스트 기법을 이용한 커널 능형 로지스틱 회귀 분류법 배깅 기법의

문장표절률: 32%

물론 여기에서도 함수 는 배깅의 경우와 마찬가지로 절단변수가 설명변수에 의존하지 않는다는 가정을 바탕으로 카플란-마이어 추정량(Kaplan-Meier estimator)을 사용하여 추정한다.

[boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

발행 : boxnwhis.kr

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다. 이론적인 생존 곡선과는 다르게

[www.boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다.

문장표절률: 0%

, , , 이 자료에 대하여 부트스트랩을 통해 번째 표본 를 추출하였다고 할 때 이 표본 안에다음과 같은 인조변수(synthetic response) 와 설명변수가 있다고 가정해보자.

문장표절률: 0%

여기에서 랜덤포레스트 기법을 적용하는 경우에는 배깅 기법의 경우와는 다르게 개의 부트스트랩 훈련자료에 대하여 개의 모든 설명변수를 포함시키지 않고 이들중 개만 선택해서 에 포함시키게 된다.

문장표절률: 0%

본 연구에서는 로 정하였으며 의 값이 자연수로 나오지 않을 시 소수점 이하 반올림을 사용하였다.

문장표절률: 0%

한가지 주의할 점은 각각의 부트스트랩 훈련자료에 포함되는 설명변수는 모두 다르게 선택해야 한다는 것이다.

문장표절률: 0%

이러한 과정을 통해 설명변수 사이의 연관성을 큰폭으로 줄일수 있으며 이를 통해 보다 더 정확한 예측값을 얻을수 있다.

문장표절률: 0%

결과적으로 랜덤포레스트 기법을 이용하여 검증자료에 대한 인조변수의 추정량을 구하는 과정은 부트스트랩 훈련자료를 추출 시 일부의 설명변수만을 선택해서 포함시킨다는 것만 차이가 있고 나머지 계산과정과 모형을 평가하는 과정은 배깅 기법의 내용과 동일하다.

문장표절률: 46%

따라서 이에 대한 자세한 설명은 생략하도록 한다. 다만, 배깅 기법을 통해서 얻어지는 추정량과 성능 평가를 위한 RMSE와의 구별을 위해 이를 아래와 같이 표기하도록 한다.

[patents.google.com] KR102175041B1 - 가상 시리얼 포트를 이용한 캡티브 포탈 주문 결제 출력 시....

발행 : 2019-07-19, 2020-11-05

결제하는 방법은 통상의 방법과 같으므로, 이에 대한 자세한 설명은 생략하도록 한다. 한편, 사업장이 제공하는 와이파이를 이용한

[clomag.co.kr] [신광섭의 데이터바로보기]데이터과학자, 네 정체가 뭐니 - CLO

발행 : clomag.co.kr

대한 지식' 역시 마찬가지다. 때문에 이에 대한 자세한 설명은 생략하도록 한다.

문장표절률: 0%

본 연구에서 제안하는 앙상블 기법을 이용한 커널 능형 중도절단 회귀분석이 실제 관심사건이 일어날 때까지 걸리는 시간에 대한 예측력의 측면에서 Huh, M.

문장표절률: 0%

(2015)에서 제안한 sub-sampling 등의 다른 방법론과 비교했을 때 전체적으로 우수한 성능을 보임을 입증하기 위해 실시한 모의실험 및 실증분석의 결과 제시, 그리고 이에 대한 해석과 평가는 6장에서 진행하도록 하겠다.

문장표절률: 0%

6 모의실험 및 실증분석 (앙상블 기법을 이용한 커널 능형 중도절단 회귀분석) 본 장에서는 2~5장에서 제안하는 앙상블 기법을 이용한 커널 능형 중도절단 회귀분석방법이 다른 방법론과 비교했을 때 전체적으로 예측력이 우수하다는 사실을 입증하기 위해 실시한 모의실험과 실증분석에 관한 내용을 설명한다.

문장표절률: 18%

우선 모의실험의 경우는 훈련자료(training data)과 검증자료(test data)를 각각 연구의 목적에 맞게 생성하여 실시하였으며, 실증분석의 경우는 실제 연구에 의하여 작성된 생존자료를 임의로 train:test=7:3 의 비율로 나누어서 진행하였다.

[juzi.tistory.com] [머신러닝 데이터 분석] Iris 품종 분류 - 티스토리

분리 1) 직접 분리 기본적으로 train:test 7:3 의 비율로 데이터를 나눈다. 비율은 조정 가능하다

[blog.naver.com] [14일차] tensorflow 심화과정

벌어질 수 있기 때문이다. 보통 train:test 7:3 의 비율로 나눈다 보다 정확하게 데이터를 나누기

문장표절률: 0%

그리고 모든 모의실험과 실증분석은 프로그램 R 4.1.1 version을 이용하여 실시하였다.

문장표절률: 44%

6.1 모의실험 모의실험에서는 평가하고자 하는 방법론의 목적에 맞게 임의로 생성한 훈련자료(training data)와 검증자료(test data)에 대해서 다음과 같은 방법론들을 비교, 분석하였다.

[www.researchgate.net] Study on Factors Influencing Self-Esteem Among Children Based on ...

서 가장 간명한 모형을 선택하였다. 훈련자료(training data)와 검증자료(test data)를 7:3 비율로 분할하여

[www.childstudies.org] 아동의 자아존중감에 영향을 미치는 요인 탐색: 빅데이터 분석을 중심으로

좋으면서 가장 간명한 모형을 선택하였다. 훈련자료(training data)와 검증자료(test data)를 7:3 비율로 분할하여

문장표절률: 0%

---------- PKR1 : Polynomial Kernel Ridge Regression with Synthetic Response PKRS1 : Polynomial Kernel Ridge Regression with Sub-sampling and Synthetic Response PKRB1 : Polynomial Kernel Ridge Regression with Bagging and Synthetic Response

문장표절률: 0%

PKRR1 : Polynomial Kernel Ridge Regression with Random Forest and Synthetic Response GKR1 : Gaussian Kernel Ridge Regression with Synthetic Response GKRS1 : Gaussian Kernel Ridge Regression with Sub-sampling and Synthetic Response GKRB1 : Gaussian Kernel Ridge Regression with Bagging and Synthetic Response

문장표절률: 0%

GKRR1 : Gaussian Kernel Ridge Regression with Random ForestandSynthetic Response PKR2 : Polynomial Kernel Ridge Regression with Generated(Original) Response PKRS2 : Polynomial Kernel Ridge Regression with Sub-sampling and Generated(Original) Response

문장표절률: 0%

PKRB2 : Polynomial Kernel Ridge Regression with Bagging and Generated(Original) Response PKRR2 : Polynomial Kernel Ridge Regression with Random Forest and Generated(Original) Response GKR2 : Gaussian Kernel Ridge Regression with Generated(Original) Response

문장표절률: 0%

GKRS2 : Gaussian Kernel Ridge Regression with Sub-sampling and Generated(Original) Response GKRB2 : Gaussian Kernel Ridge Regression with Bagging and Generated(Original) Response GKRR2 : Gaussian Kernel Ridge Regression with Random Forest and Generated(Original) Response

문장표절률: 0%

---------- 위에 제시된 16가지 방법론들 중 PKRS1, GKRS1, PKRS2, 그리고 GKRS2에서 사용되는 sub-sampling은 Huh, M. (2015) 에 의하여 제안된 방법론이다.

문장표절률: 0%

이 방법론의 원리에 대해서는 모의실험 step을 설명하면서 간단하게 언급하도록 하겠다.

문장표절률: 0%

본 모의실험에서는 다음과 같은 방법을 통해 임의의 모의실험 데이터를 생성하였다.

문장표절률: 58%

여기에서는 2장에서 소개한 카플란-마이어 추정량(Kaplan-Meier estimator)을 사용하여 계산하였다.

[boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

발행 : boxnwhis.kr

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다. 이론적인 생존 곡선과는 다르게

[www.boxnwhis.kr] 유저 생존 곡선 그리기 :: -[ ]- Box and Whisker

아래와 같다. 생존 함수 추정은 카플란-마이어 추정량(Kaplan-Meier estimator)을 이용한다.

문장표절률: 0%

그리고 반응변수는 평균이 이고 표준편차가 인 정규분포를 따르도록 생성하여 강한 비선형성을 만족하도록 하였다.

문장표절률: 0%

추가로 중도절단변수의 경우는 평균 를 모의실험 상황에 따라 적절하게 설정하여 원하는 중도절단의 비율을 나타내도록 설계하였으며, Synthetic Response와 관련된 방법론(PKR1, PKRS1, PKRB1, PKRR1, GKR1, GKRS1, GKRB1, GKRR1)에 대해서는 RMSE(Root Mean Squared Error)를 를 통해 계산하였고, Original Response와 관련된 방법론(PKR2, PKRS2, PKRB2, PKRR2, GKR2, GKRS2, GKRB2, GKRR2)에 대해서는 RMSE를 을 통해 계산하였다.

문장표절률: 64%

모의실험을 위해 생성한 임의의 훈련자료(training data)와 검증자료(test data)에 대한 설명 변수의 개수는 으로 설정하고 중도절단의 비율은 로 설정하였다.

[www.researchgate.net] Study on Factors Influencing Self-Esteem Among Children Based on ...

서 가장 간명한 모형을 선택하였다. 훈련자료(training data)와 검증자료(test data)를 7:3 비율로 분할하여

[www.childstudies.org] 아동의 자아존중감에 영향을 미치는 요인 탐색: 빅데이터 분석을 중심으로

좋으면서 가장 간명한 모형을 선택하였다. 훈련자료(training data)와 검증자료(test data)를 7:3 비율로 분할하여

문장표절률: 0%

그리고 훈련자료에 대한 관측치의 개수는 으로 설정 하였고 커널트릭 기법 적용시 2가지의 커널함수(Polynomial, Gaussian)를 적용하여 총 192가지의 상황을 가정 하였다.

문장표절률: 0%

그리고 검증자료의 경우는 관측치의 개수를 으로 고정하였다. 본 연구에서 진행한 모의실험 step은 다음과 같으며 각 방법론마다 100번의 반복을 실시하였다.

문장표절률: 0%

---------- 1) PKR1, GKR1, PKR2, GKR2 Step1) 모의실험을 하기 위한 훈련자료와 검증 자료를 각각 생성한다. Step2) 5-fold CV(cross-validation)를 통해서 최적의 능형모수(ridge parameter) 의 값을 선정한다.

문장표절률: 0%

Step3) Step2)에서 선정한 최적의 의 값과 훈련자료를 이용하여 회귀계수벡터 의 값을 구하고 이를 바탕으로 검증자료에 대한 최종적인 인조변수 추정량을 결정한 뒤 test RMSE를 계산한다.

문장표절률: 0%

2) PKRS1, GKRS1, PKRS2, GKRS2 (Sub-sampling) Step1) 모의실험을 하기 위한 훈련자료와 검증 자료를 각각 생성한다.

문장표절률: 0%

Step2) 훈련자료에 있는 관측치들 중 70%를 임의로 선택하여 새로운 훈련자료를 설정하고 나머지 30%를 이에 대한 평가자료로 사용하여 5-fold CV를 통해 최적의 의 값을 구한 뒤 test RMSE를 계산한다.

문장표절률: 30%

이 과정을 50번 반복하여 test RMSE의 값을 가장 작게 만들어주는 새로운 훈련자료와 이 훈련자료를 통해서 찾은 최적의 의 값을 최종적인 평가 기준으로 선택한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

값이 가장 작은 경우에 선택된 새로운 훈련자료와 이 훈련자료를 통해서 찾은 최적의 값을 최종적인 평가 기준으로 선택한다

문장표절률: 0%

Step3) Step2)에서 선정한 최적의 의 값과 새로운 훈련자료를 이용하여 회귀계수벡터 의 값을 구하고 이를 바탕으로 검증자료에 대한 최종적인 인조변수에 대한 sub-sampling 추정량을 결정한 뒤 test RMSE를 계산한다.

문장표절률: 0%

3) PKRB1, GKRB1, PKRB1, GKRB1 (Bagging) Step1) 모의실험을 하기 위한 훈련자료와 검증 자료를 각각 생성한다.

문장표절률: 0%

Step2) 훈련자료를 이용하여 훈련자료와 관측치의 개수가 동일한 50개의 부트스트랩 표본들을 복원추출을 통해서 생성하고 이들을 새로운 훈련자료로 설정한다.

문장표절률: 38%

각각의 부트스트랩 표본에 대한 평가 자료의 경우는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치를 모아서 설정한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

대한 새 로운 평가 자료는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치로 정한다. 물론 새로운 평가

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

대한 새 로운 평가 자료는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치로 정한다. 물론 새로운 평가

문장표절률: 0%

Step3) Step2)에서 생성한 각각의 부트스트랩 표본들에 대하여 각각의 의 값에 대한 회귀계수벡터 의 값을 구하고 이를 이용하여 test RMSE를 계산한다.

문장표절률: 0%

Step4) Step3)에서 계산한 test RMSE의 결과를 각 의 값에 따라 정리하고 평균하여 그 결과가 가장 작은 경우의 의 값을 최종적으로 결정한다.

문장표절률: 38%

만약 이에 해당하는 의 값이 여러 가지로 나타나는 경우에는 그 중 가장 큰 값을 선택하도록 한다.

[kin.naver.com] b/c분석과 npv구하는법

인 여러 가지 대안이 있는 경우에는 그 중 가장 큰 값을 갖는 대안을 선택하는 것이 바람직한

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

이에 해당하는 값이 여러 가지인 경우에는 그 중 가장 큰 값을 선택한다.

문장표절률: 19%

Step5) Step4)에서 결정한 의 값을 바탕으로 서로 다른 100개의 부트스트랩 훈련자료들과 이에 대한 평가대상인 검증 자료를 이용하여 최종적인 인조변수에 대한 배깅 추정량(bagging estimator) 을 결정한다. 이를 바탕으로 검증자료에 대한 test RMSE 또는 를 계산한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

에과 이에 따른 예측값 y-을 결정한다. 이를 바탕으로 검증자료에 대한 test MR 7T□을 계산한다. ㅡ

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

estimator) , 과 이에 따른 예측값 을 결정한다. 이를 바탕으로 검증자료에 대한 test MR 을 계산한다.

문장표절률: 0%

4) PKRR1, GKRR1, PKRR2, GKRR2 (Random forests) Step1) 모의실험을 하기 위한 훈련자료와 검증 자료를 각각 생성한다.

문장표절률: 0%

Step2) 훈련자료를 이용하여 훈련자료와 관측치의 개수가 동일한 50개의 부트스트랩 표본들을 복원추출을 통해서 생성하고 이들을 새로운 훈련자료로 설정한다.

문장표절률: 0%

이 때 각각의 표본에 대하여 개의 설명변수들을 임의로 추출하여 포함시키도록 한다.

문장표절률: 30%

설명변수들의 종류는 표본마다다르게 정한다. 각각의 부트스트랩 표본에 대한 평가 자료의 경우는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치를 모아서 설정한다.

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

대한 새 로운 평가 자료는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치로 정한다. 물론 새로운 평가

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 황성윤
발행 : 서울 : 한국외국어대학교 대학원, 2017.2

대한 새 로운 평가 자료는 복원추출 시 뽑히지 않은 번호에 해당하는 관측치로 정한다. 물론 새로운 평가

문장표절률: 0%

이러한 평가 자료에 포함시킬 설명변수들의 종류는 대응되는 부트스트랩 훈련자료의 경우와 동일하도록 설정한다.

문장표절률: 0%

Step3) Step2)에서 생성한 각각의 부트스트랩 표본들에 대하여 각각의 의 값에 대한 회귀계수벡터 의 값을 구하고 test RMSE를 계산한다.

문장표절률: 0%

Step4) Step3)에서 계산한 test RMSE의 결과를 각 의 값에 따라 정리하고 평균하여 그 결과가 가장 작은 경우의 의 값을 최종적으로 결정한다.

문장표절률: 38%

만약 이에 해당하는 의 값이 여러 가지로 나타나는 경우에는 그 중 가장 큰 값을 선택하도록 한다.

[kin.naver.com] b/c분석과 npv구하는법

인 여러 가지 대안이 있는 경우에는 그 중 가장 큰 값을 갖는 대안을 선택하는 것이 바람직한

[Copykiller] 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 = A study on efficiency of kernel ridge logistic regression classification using ensemble method

저자 : 黃星潤
발행 : 韓國外國語大學校 大學院 : 통계학과 통계학 2017. 2

이에 해당하는 값이 여러 가지인 경우에는 그 중 가장 큰 값을 선택한다.

문장표절률: 0%

Step5) Step4)에서 결정한 의 값을 토대로 서로 다른 100개의 부트스트랩 훈련자료들과 이에 대한 평가대상인 검증 자료를 이용하여 최종적인 인조변수에 대한 랜덤포레스트 추정량(random forests estimator) 을 결정한다.

문장표절률: 0%

이 때 각각의 부트스트랩 훈련자료에 대하여 개의 설명변수들을 임의로 추출해서 포함시키도록 한다.

문장표절률: 0%

설명변수들의 종류는 부트스트랩 훈련자료마다다르게 설정한다. 그리고평가 대상인 검증자료에 포함될 설명변수의 종류는 해당하는 부트스트랩 훈련자료들의 기준을 따라서 설정하도록 한다. 이를 바탕으로 검증자료에 대한 test RMSE 또는 를 계산한다.

문장표절률: 0%

---------- 위에서 설명한 step을 바탕으로 하여 각 방법론에 대해 총 192가지의 상황을 가정하고 모의실험을 진행하였으며 이를 통해 산출된 test RMSE에 대한 결과를 상자그림(boxplot)과 표로 정리하였으며 이는 〈그림 5〉~〈그림 20〉과 〈표 3〉~〈표 18〉을 통해 확인할 수 있다.

문장표절률: 0%

여기에서 상자그림의 높이와 길이는 각각 test RMSE의 평균값과 분산과 관련이 있다. 그러므로 상자그림의 높이가 낮을수록 해당하는 방법론의 예측력이 정확하다는 것을 의미하고, 상자그림의 길이가 짧을수록 해당하는 방법론의 안정성이 우수하다는 것을 의미한다.

문장표절률: 0%

이러한 사실을 바탕으로 하여 상자그림의 높이와 길이가 각각 낮고 짧게 나타나는 방법론을 예측력이 우수한 것으로 판단하면 된다.

문장표절률: 0%

추가로 〈표 3〉~〈표 18〉의 경우 각 모의실험 상황마다 가장 우수한 성능을 보인 방법론에 대하여 그 결과를 진한 글씨로 표시하여 눈에 띄도록 하였다.

문장표절률: 0%

〈그림 5 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.453 1.542 1.432 1.131 1.627 1.637 1.648 1.459 1.442 1.507 1.404 1.151 1.727 1.739 1.752 1.549 sd 0.229 0.240 0.170 0.063 0.040 0.039 0.035 0.053 0.216 0.211 0.162 0.061 0.040

문장표절률: 0%

0.041 0.036 0.054 n=100 mean 1.154 1.223 1.165 1.060 1.553 1.573 1.581 1.346 1.161 1.204 1.169 1.090 1.650 1.667 1.682 1.429 sd 0.058 0.077 0.071 0.030 0.042 0.039 0.034 0.049 0.058 0.069 0.064 0.036 0.043 0.038 0.035 0.056 n=200 mean 1.073 1.104 1.062 1.040 1.449 1.474 1.477 1.217 1.082 1.085 1.073 1.072 1.537 1.561

문장표절률: **10%**

1.573 1.297 sd 0.040 0.046 0.037 0.027 0.037 0.033 0.034 0.037 0.038 0.042 0.037 0.030 0.038 0.034 0.035 0.036 〈표 3 : , 중도절단 〉〈그림 6 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.832 1.820 1.759 1.455 1.782 1.781 1.793 1.674 1.610 1.542 1.524 1.235

[Copykiller] Habilidades, formación para el trabajo y subempleo juvenil: un enfoque de ciclo de vida

저자 : Sánchez, Alan
발행 : 2019-09

escala de autoestima 0.038 0.037 0.038 0.042 0.037 0.038 0.038 0.038 0.038

문장표절률: **0%**

1.749 1.748 1.766 1.599 sd 0.372 0.213 0.236 0.084 0.059 0.054 0.056 0.070 0.328 0.207 0.237 0.088 0.044 0.041 0.038 0.057 n=100 mean 1.506 1.547 1.518 1.381 1.740 1.736 1.746 1.593 1.270 1.224 1.278 1.155 1.692 1.685 1.706 1.497 sd 0.088 0.105 0.089 0.059 0.064 0.060 0.056 0.072 0.095 0.083 0.089 0.051 0.051 0.050

문장표절률: **0%**

0.037 0.049 n=200 mean 1.402 1.432 1.399 1.353 1.677 1.666 1.679 1.530 1.151 1.094 1.144 1.119 1.602 1.579 1.615 1.371 sd 0.069 0.065 0.071 0.057 0.061 0.064 0.056 0.059 0.055 0.045 0.055 0.042 0.045 0.044 0.037 0.056 〈표 4 : , 중도절단 〉

문장표절률: **0%**

〈그림 7 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.345 2.332 2.306 2.015 2.156 2.146 2.155 2.116 1.770 1.564 1.723 1.379 1.784 1.757 1.788 1.685 sd 0.263 0.207 0.206 0.101 0.091 0.092 0.090 0.098 0.213 0.231 0.226 0.110 0.043

문장표절률: **0%**

0.044 0.039 0.060 n=100 mean 2.129 2.112 2.132 1.962 2.156 2.120 2.142 2.082 1.500 1.253 1.483 1.278 1.763 1.697 1.753 1.621 sd 0.139 0.114 0.136 0.092 0.093 0.095 0.088 0.090 0.151 0.096 0.140 0.077 0.051 0.056 0.035 0.053 n=200 mean 1.996 2.032 1.999 1.925 2.137 2.100 2.112 2.044 1.307 1.107 1.310 1.227 1.717 1.619

문장표절률: **0%**

1.695 1.532 sd 0.109 0.101 0.111 0.093 0.108 0.100 0.090 0.094 0.078 0.053 0.077 0.057 0.069 0.061 0.040 0.049 〈표 5 : , 중도절단 〉〈그림 8 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.710 2.701 2.690 2.384 2.422 2.415 2.418 2.408 1.987 1.543 1.922 1.524

문장표절률: **9%**

1.816 1.776 1.810 1.748 sd 0.295 0.242 0.279 0.156 0.150 0.149 0.151 0.151 0.391 0.238 0.307 0.112 0.040 0.048 0.037 0.049 n=100 mean 2.491 2.521 2.488 2.336 2.426 2.405 2.415 2.396 1.667 1.260 1.646 1.430 1.815 1.716 1.794 1.708 sd 0.218 0.147 0.209 0.149 0.155 0.153 0.153 0.151 0.189 0.098 0.182 0.096 0.064 0.063

[Copykiller] Current profile and sea-bed pressure and temperature records from the northern North Sea. Challenger Cruises 84 and 85. September 1991 - November 1991

저자 : Knight, P. J.Wilkinson, M.Glorioso, P.
발행 : 1993

157 0.157 0.155 0.155 0.153 0.153 0.151 0.150 0.150 0.148

문장표절률: **0%**

0.038 0.046 n=200 mean 2.393 2.451 2.395 2.312 2.422 2.399 2.406 2.379 1.492 1.110 1.493 1.366 1.791 1.680 1.760 1.649 sd 0.168 0.139 0.170 0.152 0.151 0.148 0.150 0.149 0.098 0.045 0.097 0.074 0.062 0.080 0.040 0.064 〈표 6 : , 중도절단 〉

문장표절률: **11%**

〈그림 9 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.460 1.438 1.350 1.178 1.897 1.899 1.901 1.595 1.458 1.419 1.359 1.216 2.011 2.013 2.015 1.690 sd 0.176 0.113 0.067 0.050 0.038 0.038 0.038 0.038 0.068 0.173 0.096 0.075 0.055 0.037

[Copykiller] Work-related injuries and incentives, adequacy, and optimality in workers' compensation.

저자 : Bronchetti, Erin Todd

037) (0.0 050 038 0.050 0.038 0.038 0.038 0.038 0.039) Female: Never married -0

문장표절률: **9%**

0.038 0.037 0.068 n=100 mean 1.403 1.398 1.312 1.138 1.887 1.891 1.893 1.457 1.381 1.357 1.290 1.178 2.001 2.005 2.007 1.543 sd 0.145 0.071 0.059 0.033 0.037 0.038 0.037 0.057 0.150 0.070 0.061 0.038 0.037 0.038 0.037 0.068 n=200 mean 1.245 1.302 1.234 1.123 1.863 1.872 1.877 1.309 1.232 1.252 1.207 1.165 1.976 1.986

[stackoverflow.com] In ggplot, how can I plot time series data like this picture? - Stack ...

발행 : stackoverflow.com

038, 0.04, 0.039 0.038 0.037 0.038 0.037 0.035, 0.035, 0.037

[Copykiller] 연구보고서 20-14 조세 · 재정 정책과 기업의 고용조정에 관한 연구: 청년고용 임금보조금과 세액공제를 중심으로

저자 : 김문정 오종현 조원기
발행 : 2021-03-31

038 0.037 0.037 0.038 0.037 0.038 0.037 0.037 0.039) ln(매

문장표절률: **10%**

1.991 1.398 sd 0.055 0.062 0.048 0.028 0.039 0.038 0.038 0.044 0.056 0.067 0.047 0.030 0.038 0.038 0.038 0.043 〈표 7 : , 중도절단 〉〈그림 10 : , 중도절단 〉PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.897 1.717 1.700 1.518 2.022 2.023 2.024 1.808 1.629 1.423 1.466 1.280

[Copykiller] Evaluating five different Loci (rbcL, rpoB, rpoC1, matK and ITS) for DNA Barcoding of Indian Orchids

저자 : Parveen, Iffat, Singh, Hemant K., Malik, Saloni, Raghuvanshi, Saurabh, Babbar, Shashi B.
발행 : 2017-04-21

073 0.104 0.040 0.030 0.038 0.038 0.038 0.107 0.097 0.075 ..... 073 0.104 0.040 0.030 0.038 0.038 0.038 0.107 0.097 0.075

[Copykiller] Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids.

저자 : Iffat Parveen, Hemant Kumar Singh, Saloni Malik, Saurabh Raghuvanshi, Shashi B Babbar
발행 : 2017

073 0.104 0.040 0.030 0.038 0.038 0.038 0.107 0.097 0.075 ..... 073 0.104 0.040 0.030 0.038 0.038 0.038 0.107 0.097 0.075

문장표절률: **10%**

2.013 2.013 2.015 1.738 sd 0.371 0.088 0.096 0.065 0.056 0.055 0.056 0.071 0.229 0.088 0.103 0.063 0.038 0.037 0.037 0.063 n=100 mean 1.802 1.711 1.704 1.471 2.014 2.015 2.018 1.706 1.530 1.365 1.421 1.229 2.003 2.005 2.008 1.611 sd 0.167 0.085 0.105 0.057 0.056 0.056 0.056 0.070 0.125 0.064 0.097 0.048 0.038 0.037

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

0.056 9 Tochigi(Utsunomiya) 0.057 0.056 0.056 0.056 0.055 0.056 0.056 ..... Shinjuku) 0.057 0.057 0.057 0.056 0.056 0.056 0.056 0.057 0.057

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

0.044 13 Tokyo(Shinjuku) 0.057 0.056 0.056 0.056 0.056 0.056 0.056 0.056

문장표절률: **24%**

0.037 0.050 n=200 mean 1.659 1.634 1.633 1.454 1.996 2.000 2.006 1.654 1.385 1.250 1.337 1.209 1.983 1.986 1.994 1.470 sd 0.093 0.070 0.088 0.056 0.057 0.056 0.056 0.076 0.097 0.063 0.079 0.037 0.038 0.038 0.038 0.074 〈표 8 : , 중도절단 〉

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

038 0.037 0.037 0.037 0.038 0.038 0.038 0.038 0.038 0.038 ..... 037 0.037 0.037 0.037 0.037 0.038 0.038 0.038 0.038 0.038 0.038 0.038

[usermanual.wiki] Poisson Superfish Los Alamos Manual

037 0 037 0 037 0.037 0.038 0.038 0 038 0 039 0.039 (),~39 by

문장표절률: **0%**

〈그림 11 : , 중도절단 〉PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.592 2.278 2.366 2.148 2.373 2.372 2.374 2.296 1.974 1.436 1.734 1.427 2.016 2.014 2.018 1.838 sd 0.379 0.119 0.161 0.125 0.107 0.107 0.107 0.117 0.254 0.086 0.174 0.109 0.038

문장표절률: **10%**

0.038 0.038 0.066 n=100 mean 2.599 2.316 2.429 2.093 2.369 2.368 2.371 2.253 1.932 1.372 1.738 1.341 2.009 2.006 2.012 1.747 sd 0.282 0.117 0.152 0.110 0.107 0.107 0.107 0.115 0.228 0.075 0.162 0.068 0.038 0.038 0.037 0.066 n=200 mean 2.379 2.269 2.342 2.071 2.362 2.360 2.364 2.193 1.687 1.254 1.616 1.311 1.993 1.988

[Copykiller] Chapter 12.B Skilled and Unskilled Labor Data

저자 : Betina V. Dimaranan, Badri Narayanan G

107 0.107 0.107 0.107 0.107 0.107 0.115 0.124 0.115 0.119

[Copykiller] Current profile and sea-bed pressure and temperature records from the northern North Sea. Challenger Cruises 84 and 85. September 1991 - November 1991

저자 : Knight, P. J.Wilkinson, M.Glorioso, P.
발행 : 1993

114 0.111 0.110 0.110 0.107 0.107 0.107 0.103 0.101 0.100

문장표절률: **10%**

2.001 1.654 sd 0.150 0.118 0.135 0.112 0.108 0.108 0.107 0.121 0.139 0.048 0.118 0.054 0.038 0.038 0.038 0.047 〈표 9 : , 중도절단 〉〈그림 12 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 3.016 2.712 2.809 2.572 2.651 2.650 2.651 2.630 2.218 1.437 1.963 1.585

[www.ide.go.jp] Impact of Diagonal Cumulation Rule on FTA Utilization ...

053 0.053 0.054 0.054 0.038 0.038 0.038 0.038]

문장표절률: **0%**

2.018 2.015 2.019 1.913 sd 0.362 0.170 0.241 0.164 0.168 0.168 0.168 0.169 0.341 0.087 0.216 0.120 0.038 0.038 0.038 0.052 n=100 mean 3.043 2.779 2.906 2.531 2.651 2.649 2.651 2.616 2.209 1.375 2.015 1.506 2.014 2.008 2.016 1.860 sd 0.296 0.169 0.250 0.163 0.168 0.169 0.168 0.168 0.309 0.074 0.232 0.080 0.037 0.038

문장표절률: **0%**

0.037 0.055 n=200 mean 2.825 2.757 2.794 2.511 2.648 2.643 2.648 2.595 1.939 1.256 1.867 1.468 2.006 1.990 2.009 1.788 sd 0.204 0.150 0.190 0.166 0.168 0.168 0.168 0.170 0.176 0.059 0.147 0.074 0.039 0.038 0.039 0.077 〈표 10 : , 중도절단 〉

문장표절률: **11%**

〈그림 13 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.404 1.401 1.387 1.263 2.039 2.039 2.039 1.926 1.443 1.426 1.430 1.268 2.159 2.159 2.159 2.045 sd 0.096 0.074 0.059 0.047 0.036 0.036 0.036 0.037 0.100 0.074 0.079 0.051 0.036

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

036 0.036 0.036 0.036 0.036 0.036 0.037 0.039 0.0306~0.0943

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

036 0.035 0.036 0.036 0.036 0.036 0.037 0.0306~0.0943 6.5

문장표절률: **17%**

0.036 0.036 0.038 n=100 mean 1.332 1.308 1.285 1.176 2.037 2.038 2.038 1.831 1.342 1.300 1.299 1.207 2.157 2.158 2.158 1.949 sd 0.077 0.044 0.045 0.033 0.036 0.036 0.036 0.038 0.088 0.051 0.050 0.038 0.036 0.036 0.036 0.039 n=200 mean 1.304 1.285 1.242 1.152 2.034 2.036 2.036 1.683 1.280 1.248 1.227 1.187 2.154 2.155

[usermanual.wiki] Poisson Superfish Los Alamos Manual

0.00000 0,00000 (gauss) 0 038 0.036 0.036 0.036 0.036 0.036 0.036 0.036

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

035 0.035 0.037 0.038 0.036 0.036 0.036 0.036 0.035 0.036 0.035

문장표절률: **10%**

2.156 1.796 sd 0.113 0.040 0.033 0.027 0.036 0.036 0.036 0.037 0.098 0.040 0.036 0.034 0.036 0.037 0.036 0.038 〈표 11 : , 중도절단 〉〈그림 14 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.783 1.707 1.747 1.657 2.178 2.178 2.178 2.094 1.568 1.436 1.534 1.342

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

036 0.036 0.036 0.036 0.036 0.036 0.037 0.039 0.0306~0.0943

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

036 0.035 0.036 0.036 0.036 0.036 0.037 0.0306~0.0943 6.5

문장표절률: **17%**

2.159 2.159 2.159 2.058 sd 0.113 0.071 0.072 0.069 0.055 0.055 0.055 0.056 0.123 0.067 0.085 0.072 0.036 0.036 0.036 0.039 n=100 mean 1.776 1.674 1.700 1.594 2.177 2.177 2.178 2.025 1.487 1.302 1.416 1.287 2.158 2.158 2.158 1.973 sd 0.104 0.061 0.064 0.059 0.055 0.055 0.055 0.057 0.089 0.054 0.065 0.053 0.036 0.036

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

0.064 37 Kagawa(Takamatsu) 0.055 0.055 0.055 0.056 0.056 0.055 0.055

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture（14:00 ...

발행 : radioactivity.nsr.go.jp

054 0.055 0.054 0.055 0.055 0.055 0.056 0.056 0.055 0.055

문장표절률: **12%**

0.036 0.040 n=200 mean 1.785 1.676 1.700 1.555 2.175 2.175 2.176 1.927 1.45 1 1.252 1.374 1.253 2.155 2.156 2.157 1.842 sd 0.148 0.064 0.073 0.055 0.055 0.054 0.055 0.060 0.086 0.039 0.063 0.045 0.036 0.036 0.036 0.042 〈표 12 : , 중도절단 〉

[stackoverflow.com] In ggplot, how can I plot time series data like this pictur e? - Stack ...

발행 : stackoverflow.com

058, 0.056, 0.057, 0.055 0.055 0.054 0.055 0.053 0.053 0.051

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefec ture（14:00 ...

발행 : radioactivity.nsr.go.jp

060 0.056 0.055 0.055 0.055 0.054 0.055 0.055 0.055 0.028

문장표절률: **0%**

〈그림 15 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 G KRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mea n 2.429 2.296 2.371 2.313 2.536 2.536 2.536 2.500 1.771 1.446 1.716 1.502 2.15 9 2.159 2.159 2.086 sd 0.134 0.119 0.116 0.115 0.103 0.103 0.103 0.105 0.139 0.082 0.103 0.106 0.036

문장표절률: **9%**

0.036 0.036 0.041 n=100 mean 2.506 2.310 2.393 2.259 2.535 2.535 2.535 2.47 4 1.787 1.312 1.656 1.446 2.158 2.158 2.159 2.027 sd 0.188 0.108 0.130 0.114 0.103 0.103 0.103 0.106 0.150 0.056 0.110 0.081 0.036 0.036 0.036 0.043 n= 200 mean 2.546 2.342 2.431 2.216 2.534 2.534 2.535 2.435 1.792 1.253 1.664 1. 391 2.156 2.156

[Copykiller] 부가가치세 유효 세부담 변화 분석과 정책방향

저자 : 박명호
발행 : 2014

및 이미용용품 칫솔 0.103 0.103 0.103 0.103 0.106 0.106 0.106 치약 0 ..... 0.108 종교용품 0.103 0.103 0.103 0.103 0.106 0.106 0.106

[Copykiller] Evaluating five different Loci (rbcL, rpoB, rpoC1, matK and ITS) for DNA Barcoding of Indian Orchids

저자 : Parveen, Iffat, Singh, Hemant K., Malik, Saloni, Raghuvanshi, Saurabh, Babbar, Shashi B.
발행 : 2017-04-21

103 0.103 0.103 0.103 0.103 0.103 0.106 0.114 0.139 0.130

문장표절률: **10%**

2.157 1.945 sd 0.246 0.116 0.142 0.112 0.104 0.104 0.103 0.106 0.155 0.043 0.123 0.066 0.037 0.036 0.036 0.050 〈표 13 : , 중도절단 〉〈그림 16 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.884 2.757 2.819 2 .788 2.851 2.851 2.851 2.840 1.959 1.454 1.873 1.679

[Copykiller] 부가가치세 유효 세부담 변화 분석과 정책방향

저자 : 박명호
발행 : 2014

106 기타운송기구구입 오토바이 0.103 0.104 0.104 0.103 0.106 0.107 0.108 자전거등 0

문장표절률: **9%**

2.159 2.159 2.159 2.112 sd 0.217 0.172 0.193 0.179 0.179 0.179 0.179 0.180 0. 156 0.074 0.098 0.117 0.036 0.036 0.036 0.039 n=100 mean 2.992 2.814 2.88 0 2.747 2.851 2.851 2.851 2.833 2.040 1.306 1.876 1.636 2.159 2.158 2.159 2.0 84 sd 0.226 0.165 0.195 0.174 0.179 0.179 0.179 0.178 0.228 0.055 0.141 0.1 06 0.036 0.036

[fraser.stlouisfed.org] Full text of Wholesale Prices : Wholesale Prices, 1890 t o 1925 ...

발행 : fraser.stlouisfed.org

179 0 179.4 179 0 179 0 179 0 179 0 179.0 8.453 9.920 8 ..... 5 147.5 966 179 0 179 0 179 0 179 0 179 0 179 0 8.320 8.320 8 ..... 5 147.5 696 179 0 179 0 1 79 0 179 0 179 0 179 0 179 0 8.320 8.320 8 ..... 147.5 147.5 179 0 179 0 179 0 179 0 179 0 179 0 8.320 8.320 8

[fraser.stlouisfed.org] Full text of Wholesale Prices : Wholesale Prices, 1890 t o 1926 ...

발행 : fraser.stlouisfed.org

696 .696 .696 .696 179 0 179 0 179 0 179 0 179 0 179 0 179 0 8. 320 8. 320 8 ..... 696 .696 .696 .696 179 0 179 0 179 0 179 0 179 0 179 0 8. 320 8. 320 8 ..... 696 .696 .696 .696 179 0 179 0 179 0 179 0 179 0 179 0 179 0 8. 320 8 . 320 8

문장표절률: **0%**

0.036 0.040 n=200 mean 3.033 2.868 2.941 2.710 2.850 2.850 2.850 2.819 2.0 50 1.251 1.906 1.576 2.157 2.156 2.158 2.036 sd 0.201 0.168 0.185 0.177 0.18 0 0.179 0.179 0.179 0.155 0.042 0.127 0.086 0.036 0.036 0.036 0.052 〈표 14 : , 중도절단 〉

문장표절률: **11%**

〈그림 17 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 G KRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mea n 1.428 1.419 1.403 1.285 2.145 2.145 2.146 2.022 1.460 1.450 1.448 1.294 2.2 67 2.267 2.267 2.144 sd 0.079 0.075 0.070 0.051 0.040 0.040 0.040 0.041 0.0 85 0.089 0.086 0.054 0.040

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefec ture（14:00 ...

발행 : radioactivity.nsr.go.jp

Saga) 0.039 0.040 0.040 0.040 0.040 0.040 0.041 0.041 0.041 0.040

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefec ture（14:00 ...

발행 : radioactivity.nsr.go.jp

040 0.040 0.040 0.040 0.040 0.040 0.040 0.041 0.041 0.037~0.086

문장표절률: **19%**

0.040 0.040 0.041 n=100 mean 1.348 1.331 1.326 1.217 2.145 2.145 2.145 1.923 1.384 1.339 1.352 1.254 2.267 2.267 2.267 2.042 sd 0.097 0.064 0.046 0.038 0.040 0.040 0.040 0.039 0.125 0.059 0.061 0.043 0.040 0.040 0.040 0.039 n=200 mean 1.275 1.255 1.249 1.206 2.145 2.145 2.145 1.768 1.287 1.251 1.263 1.248 2.267 2.267

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

051 0.048 0.042 0.040 0.040 0.040 0.039 0.040 0.040 0.040 0.039 0.039 0.037~0.086

[link.springer.com] Decentralisation and political inequality: a comparative analysis of ...

저자 : Birte Gundelach, Matthias Fatke
발행 : 2019/09/23

Gender 0.039 0.005 0.040 0.040 0.040 0.039 0.03) (0.03) (0.03

문장표절률: **24%**

2.267 1.883 sd 0.060 0.038 0.035 0.032 0.040 0.040 0.040 0.039 0.063 0.045 0.040 0.035 0.040 0.040 0.040 0.039 〈표 15 : , 중도절단 〉〈그림 18 : , 중도절단 〉PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 1.821 1.751 1.783 1.681 2.278 2.278 2.278 2.185 1.582 1.469 1.542 1.355

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

051 0.048 0.042 0.040 0.040 0.040 0.039 0.040 0.040 0.040 0.039 0.039 0.037~0.086

[link.springer.com] Decentralisation and political inequality: a comparative analysis of ...

저자 : Birte Gundelach, Matthias Fatke
발행 : 2019/09/23

Gender 0.039 0.005 0.040 0.040 0.040 0.039 (0.03) (0.03) (0.03

문장표절률: **9%**

2.267 2.267 2.267 2.156 sd 0.108 0.088 0.090 0.077 0.065 0.065 0.065 0.068 0.112 0.106 0.089 0.072 0.040 0.040 0.040 0.041 n=100 mean 1.782 1.677 1.718 1.625 2.278 2.278 2.278 2.112 1.505 1.360 1.470 1.330 2.267 2.267 2.267 2.066 sd 0.153 0.074 0.076 0.075 0.065 0.065 0.065 0.069 0.127 0.058 0.064 0.052 0.040 0.040

[Copykiller] Evaluating five different Loci (rbcL, rpoB, rpoC1, matK and ITS) for DNA Barcoding of Indian Orchids

저자 : Parveen, Iffat, Singh, Hemant K., Malik, Saloni, Raghuvanshi, Saurabh, Babbar, Shashi B.
발행 : 2017-04-21

070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077

[Copykiller] Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids.

저자 : Iffat Parveen, Hemant Kumar Singh, Saloni Malik, Saurabh Raghuvanshi, Shashi B Babbar
발행 : 2017

070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.106 0.096 0.077

문장표절률: **13%**

0.040 0.040 n=200 mean 1.717 1.643 1.678 1.598 2.278 2.278 2.278 2.005 1.415 1.253 1.375 1.307 2.267 2.267 2.267 1.929 sd 0.103 0.073 0.073 0.069 0.065 0.065 0.065 0.070 0.087 0.041 0.054 0.039 0.040 0.040 0.040 0.040 〈표 16 : , 중도절단 〉

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

039 0.039 0.040 0.039 0.040 0.040 0.040 0.040 0.035 0.074 24 ..... 039 0.039 0.039 0.039 0.040 0.040 0.040 0.040

[radioactivity.nsr.go.jp] Reading of environmental radioactivity level by prefecture (14:00 ...

발행 : radioactivity.nsr.go.jp

039 0.039 0.039 0.039 0.040 0.040 0.040 0.041 0.041 0.041 ..... 040 0.039 0.040 0.039 0.040 0.040 0.040 0.040 0.040 0.040 0

문장표절률: **11%**

〈그림 19 : , 중도절단 〉PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.469 2.362 2.437 2.379 2.663 2.663 2.663 2.623 1.809 1.513 1.760 1.517 2.267 2.267 2.267 2.185 sd 0.168 0.132 0.130 0.130 0.123 0.123 0.123 0.124 0.142 0.111 0.106 0.106 0.040

[www.researchgate.net] (PDF) Innovation, innovation strategy and internationalization of ...

발행 : www.researchgate.net

863*** 0.866*** (0.123 0.123 0.123 0.123 0.124 0.123 0.124 0.121

[Copykiller] Three essays in corporate governance (다운로드)

저자 : Shinwoo Kang. Three essays in corporate governance

Rsquare 0.123 0.123 0.123 0.123 0.123 0.124 0.298 0.299 0.299

문장표절률: **0%**

0.040 0.040 0.043 n=100 mean 2.439 2.335 2.405 2.332 2.663 2.663 2.663 2.590 1.686 1.375 1.650 1.474 2.267 2.267 2.267 2.117 sd 0.142 0.128 0.135 0.131 0.123 0.122 0.123 0.129 0.105 0.063 0.081 0.067 0.040 0.040 0.040 0.047 n=200 mean 2.476 2.351 2.410 2.301 2.663 2.663 2.663 2.548 1.659 1.253 1.581 1.430 2.267 2.267

문장표절률: **17%**

2.267 2.031 sd 0.158 0.120 0.126 0.123 0.123 0.123 0.123 0.126 0.098 0.038 0.070 0.057 0.040 0.040 0.040 0.049 〈표 17 : , 중도절단 〉〈그림 20 : , 중도절단 〉 PKR 1 PKRS1 PKRB1 PKRR1 GKR 1 GKRS1 GKRB1 GKRR1 PKR 2 PKRS2 PKRB2 PKRR2 GKR 2 GKRS2 GKRB2 GKRR2 n=50 mean 2.936 2.835 2.907 2.890 2.980 2.980 2.980 2.970 1.985 1.571 1.948 1.729

[www.researchgate.net] (PDF) Innovation, innovation strategy and internationalization of ...

발행 : www.researchgate.net

865*** 0.863*** 0.866 0.123 0.123 0.123 0.123 0.124) (0.123 0.124

[archive.org] Full text of "IS 1448-47: Methods of Test for Petroleum and its ...

발행 : archive.org

123 1-8 19 1 0 123 0 123 0 123 0 123 0 122 0-122 0-122

문장표절률: **0%**

2.267 2.267 2.267 2.216 sd 0.202 0.180 0.191 0.188 0.178 0.179 0.179 0.178 0.147 0.083 0.103 0.129 0.040 0.040 0.040 0.042 n=100 mean 2.956 2.852 2.906 2.842 2.980 2.980 2.980 2.959 1.907 1.377 1.850 1.664 2.267 2.267 2.267 2.175 sd 0.180 0.171 0.172 0.177 0.179 0.178 0.178 0.179 0.133 0.054 0.107 0.106 0.040 0.040

문장표절률: **29%**

0.040 0.049 n=200 mean 2.996 2.902 2.939 2.816 2.980 2.980 2.980 2.946 1.910 1.254 1.821 1.628 2.267 2.267 2.267 2.131 sd 0.185 0.172 0.177 0.177 0.179 0.179 0.179 0.177 0.125 0.039 0.088 0.072 0.040 0.040 0.040 0.040 0.051 〈표 18 : , 중도절단 〉

[Copykiller] Evaluating five different Loci (rbcL, rpoB, rpoC1, matK and ITS) for DNA Barcoding of Indian Orchids

저자 : Parveen, Iffat, Singh, Hemant K., Malik, Saloni, Raghuvanshi, Saurabh, Babbar, Shashi B.
발행 : 2017-04-21

070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077

[Copykiller] Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids.

저자 : Iffat Parveen, Hemant Kumar Singh, Saloni Malik, Saurabh Raghuvanshi, Shashi B Babbar
발행 : 2017

070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077 ..... 070 0.062 0.080 0.072 0.040 0.040 0.040 0.040 0.106 0.096 0.077

문장표절률: **0%**

〈그림 5〉~〈그림 20〉, 그리고 〈표 3〉~〈표 18〉을 통해 알 수 있듯이 총 192가지 상황을 가정하고 모의실험을 진행한 결과 전체적으로 커널 능형 중도절단 회귀분석방법에 랜덤포레스트 기법을 적용한 경우 예측력이 좋아진다는 것을 확인할 수 있다.

문장표절률: **0%**

물론 설명변수의 개수가 많고 중도절단의 비율이 큰 상황에서는 분포의 변동성이 커지기 때문에 랜덤포레스트 기법에의한 성능의 향상이 일어나지 않는 경우도 존재할 수 있다.

문장표절률: **0%**

하지만 전체적으로 살펴본다면 다양한 모의실험 상황을 가정 했음에도 불구하고 배깅이나 랜덤포레스트 등의 앙상블 기법을 적용했을 때 확실히 예측력이 좋아진다는 사실을 알 수 있다.

문장표절률: **0%**

이에 따라 본 모의실험을 통해 앙상블 기법을 적용한 커널 능형 중도절단 회귀분석방법이 다른 방법론과 비교했을 때 예측력이 우수하다는 사실을 입증할 수 있었다.

문장표절률: **20%**

6.2 실증분석 실증분석에서는 총 5가지의 실제 중도절단이 포함된 데이터를 임의로 train:test=7:3 의 비율로 분할한 뒤 6.1절에서 소개한 step을 통하여 방법론들의 예측력을 비교하기 위한 분석을 진행하였다.

[juzi.tistory.com] [머신러닝 데이터 분석] Iris 품종 분류 - 티스토리

분리 1) 직접 분리 기본적으로 train:test 7:3 의 비율로 데이터를 나눈다. 비율은 조정 가능하다

[blog.naver.com] [14일차] tensorflow 심화과정

벌어질 수 있기 때문이다. 보통 train:test 7:3 의 비율로 나눈다 보다 정확하게 데이터를 나누기

문장표절률: **0%**

다만 3.4절에서도 언급했듯이 실제 중도절단이 포함된 데이터에서는 모든 관측치에 대하여 실제 관심사건이 일어나기까지 걸린 시간을 알 수 없다는 점을 고려하여 평가기준을 인조변수로 두는 방법론에 대해서만 분석을 진행 하였다.

문장표절률: 0%

그리고 6.1절의 모의실험 결과를 통해서도 알 수 있듯이 커널트릭 기법 적용시 다항커널(Polynomial kernel)을 사용하는 경우에는 변동성이 매우 크고 이상치도 자주 발생 하게되어 분석에 대한 안정성이 떨어지는 단점이 있다.

문장표절률: 0%

그렇기 때문에 다항커널을 이용하여 실제 데이터를 분석하는 경우에는 역행렬(inverse matrix)을 계산시 나오게 되는 행렬식(determinant)의 값이 은 아니지만 에 매우 가까운 양의 실수인 경우가 빈번하게 발생하기 때문에 알고리즘이 중단되는 경우가 종종 나타난다.

문장표절률: 0%

이에 따라 실증분석에서는 모든 데이터에 대해 비교적 좋은 유연성(flexibility)과 안정성(stability)을 보이는가우시안 커널(Gaussian kernel)에 한하여 분석을 진행하였다.

문장표절률: 0%

이러한 원인에 따라 실증분석에서는 16가지 방법론들 중 GKR1, GKRS1, GKRB1, 그리고 GKRR1 이렇게 4가지 방법론들에 대하여 비교, 분석을 실시하였다.

문장표절률: 0%

그리고 각 방법론에 대해서는 100번의 반복을 실시하였다. 본 실증분석에서 사용된 모든 데이터는 프로그램 R 4.1.1 version의 생존분석 관련 package인 survival, KMsurv, survMisc, 그리고 survMiner에 내장되어 있다.

문장표절률: 0%

#1) UIS Data 본 데이터는 미국의 University of Massachusetts에서 1989년부터 1994년까지 약 5년간 협동연구로 진행한 AIDS Research Unit (UMARU) IMPACTStudy (UIS)에 대한 결과물이다.

문장표절률: 0%

주된 목적은 환자의 약물남용(drug abuse)을 조사하는 것이며 예측변수는 환자가 유혹을 참지 못하고 다시 약물에 손을 대기까지 걸린 시간(days)이다.

문장표절률: 0%

본 연구에서는 2가지의 다른 치료 프로그램을 각각 A site와 B site에서 실시하였으며, 이 중 A site에 해당하는 경우에 한하여 실증분석을 실시하였다.

문장표절률: 0%

관측치의 개수는 결측치와 이상치를 제외하여 총 398개이며 사용한 설명변수는 총 8개이다.

문장표절률: 0%

그리고 중도절단의 비율은 약 20%이다. 본 데이터에 대한 자세한 설명은 Hosmer et al. (2008)을 참고하기 바란다.

문장표절률: 0%

〈그림 21 : UIS data 실증분석〉 실증분석 결과 다른 방법론들에 비해 랜덤포레스트를 적용하였을 때 test RMSE의 평균값이 작게 산출됨을 확인할수 있었다.

문장표절률: 0%

이를 토대로 앙상블 기법을 적용한 커널 능형 중도절단 회귀분석방법이 다른 방법론들과 비교했을 때 예측력이 좋다고 판단할 수 있겠다.

문장표절률: 0%

#2) PBC Data 본 데이터는 미국의 Mayo Clinic에서 1974년부터 1984년까지 약 10년간 수행된 원발성 담즙성 간경화증(primary biliary cholangitis) 환자에 대한 연구를 통해서 나온 결과물이다.

문장표절률: 0%

주된 목적은 위약(placebo)과 D-penicillamine의 효능을 비교하는 것이다. 예측변수는 환자의 생존 시간이며 관측치의 개수는 결측치를 제외하여 총 276개이다.

문장표절률: 0%

그리고 사용한 설명변수는 총 17개이며 중도절단된 비율은 약 50%이다. 본 데이터에 대한 자세한 설명은 Therneau, T. and Grambsch, P. (2000)을 참조하기 바란다.

문장표절률: 0%

〈그림 22 : PBC data 실증분석〉 실증분석 결과 랜덤포레스트를 적용한 경우에 대해 성능의 향상이 크게 일어나지는 않았다.

문장표절률: 0%

이는 설명변수의 개수가 17개로 많으며 중도절단의 비율도 약 50%로서 매우 큰 편이기 때문에 복잡성이 커짐에 따른 결과라고 판단된다.

문장표절률: 0%

하지만 이러한 조건에도 불구하고 랜덤포레스트를 적용한 커널 능형 중도절단 회귀분석을 적용시 다른 방법론들보다 test RMSE의 평균값이 작게 산출되었기 때문에 본 연구에서 주장하고자 하는 바를 입증하는 데는 무리가 없다고 생각한다.

문장표절률: 0%

#3) Cancer Data 본 데이터는 북아메리카 지역의 암(cancer) 전문가 네트워크로 구성된 North Central Cancer Treatment Group에서 실시한 폐암(lung cancer) 환자에 대한 연구를 통해서 나온 결과물이다.

문장표절률: 0%

폐암 환자의 생존시간을 예측하는 것이 주된 목적이며 관측치의 개수는 결측치를 제외하여 총 167개이다. 그리고 사용한 설명변수는 총 7개이며 중도절단된 비율은 약 30%이다.

문장표절률: 0%

〈그림 23 : Cancer data 실증분석〉 실증분석 결과 UIS data의 경우와 마찬가지로 다른 방법론들에 비해 랜덤포레스트를 적용하였을 때 test RMSE의 평균값이 작게 산출됨을 확인할수 있었다.

문장표절률: 0%

이를 토대로 앙상블 기법을 적용한 커널 능형 중도절단 회귀분석방법이 다른 방법론들과 비교했을 때 예측력이 좋다고 판단할 수 있겠다.

문장표절률: 0%

#4) Retinopathy Data 본 데이터는 당뇨병성 망막병증(diabetic retinopathy)을 지연시키는 치료방법으로 레이저 응고법(laser coagulation)의 효과를 검증하는 연구를 통해서 나온 결과물이다.

문장표절률: 0%

시력을 잃을 때까지 걸리는 시간을 예측하는 것이 주된 목적이며 관측치의 개수는 총 394개이다.

문장표절률: 0%

사용한 설명변수는 총 6개이며 중도절단의 비율은 약 60%이다. 데이터에 포함된 변수는 다음과 같다.

문장표절률: 0%

〈그림 24 : Retinopathy data 실증분석〉 실증분석 결과 PBC data의 경우와 마찬가지로 랜덤포레스트를 적용한 경우에 대해 성능의 향상이 크게 일어나지는 않았다.

문장표절률: 0%

이는 중도절단의 비율이 약 60%로서 매우 큰 편이기 때문에 복잡성이 커짐에 따른 결과라고 판단된다.

문장표절률: 0%

하지만 이러한 조건에도 불구하고 랜덤포레스트를 적용한 커널 능형 중도절단 회귀분석을 적용시 다른 방법론들보다 test RMSE의 평균값이 작게 산출되었기 때문에 본 연구에서 주장하고자 하는 바를 입증하는 데는 무리가 없다고 생각한다.

문장표절률: 0%

#5) Bfeed Data 본 데이터는 태아를 출산한 산모의 모유수유 기간(duration of breast feeding)에 대한 연구를 통해서나온 결과물이다.

문장표절률: 0%

모유수유 기간을 예측하는 것이 주된 목적이며 관측치의 개수는 총 927개이다. 사용한 설명변수는 총 8개이며 중도절단의 비율은 약 4%이다.

문장표절률: 0%

본 데이터에 대한 자세한 사항은 Klein and Moeschberger (1997)를 참조하기 바란다. 데이터에 포함된 변수는 다음과 같다.

문장표절률: 0%

〈그림 25 : Bfeed data 실증분석〉 실증분석 결과 확실히 배깅이나 랜덤포레스트 등의 앙상블 기법을 적용했을 때 test RMSE에 대한 평균값과 분산이 작아진다는 것을 확인할수 있었다.

문장표절률: 0%

물론 중도절단의 비율이 약 4%이고이는 그렇게 크지 않은 비율이기 때문에 중도절단의 비율이 큰 데이터에 비해 앙상블 기법을 적용한 방법론에 대한 성능의 향상이 많이 일어났다고 볼 수도 있다.

문장표절률: 0%

하지만 결과적으로는 앙상블 기법을 적용한 커널 능형 중도절단 회귀분석방법이 다른 방법론들에 비해 예측력이 우수하다는 사실을 입증하기에는 충분하다고 판단된다.

문장표절률: 0%

#6) 실증분석 결과 정리 지금까지 실시한 실증분석에서 산출된 test RMSE에 대한 결과를 종합하면 〈표 19〉와 같이 정리할 수 있다.

문장표절률: 0%

data GKR1 GKRS1 GKRB1 GKRR1 Censoring rate number of explanatory variables number of observations UIS

문장표절률: 0%

mean 186.132 187.691 187.857 154.319 20% 8 398 sd 26.033 25.899 25.836 28.474 PBC mean 4300.305 4300.305 4300.305 4236.192 50% 17 276 sd 1135.303 1135.303 1135.303 1145.583 Cancer mean 509.320 509.327 509.331 465.783 30% 7 167 sd 72.125 72.137 72.138 74.089 Retinopathy mean 29.230 29.117

문장표절률: 0%

28.859 28.145 60% 6 394 sd 8.663 8.757 8.704 8.788 Bfeed mean 16.192 15.639 15.502 15.366 4% 8 927 sd 1.069 0.696 0.677 0.664 〈표 19 : 실증분석 결과 정리〉 결과적으로 실증분석에서 사용된 데이터마다 중도절단의 비율과 사용한 설명변수의 개수가 다르기 때문에 성능의 향상이 일어나는 정도에 차이가 있는 것은 분명한 사실이다.

문장표절률: 0%

하지만 그럼에도 불구하고 전체적으로 살펴봤을 경우에는 랜덤포레스트 기법을 적용했을 때 관심사건이 일어나기까지 걸리는 시간에 대한 예측력이 더 좋아진다는 것을 확인할 수 있었다.

문장표절률: 0%

이에 따라 본 실증분석을 통해서 앙상블 기법을 적용한 커널 능형 중도절단 회귀분석방법이 다른 방법론들과 비교했을 때 전체적으로 예측력이 우수하다는 사실을 입증할 수 있었다. 7 결론

참고문헌

참고문헌 Buckley, J. and James, I. (1979). Linear regression with censored data, Biometrika, 66, 429-436. Koul, H., Susarla, V., Van Ryzin, J. (1981). Regression analysis with randomly right censored data. Annals of Statistics. 9, 1276-1288. Beran, R. (1981). Non-parametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley. Suárez, R.P., Abad, R.C., and Fernández, J.M.V. (2021). Bootstrap Selector for the Smoothing Parameter of Beran's Estimator. Engineering Proceedings. Geerdens, C., Acar, E.F. and Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. Biostatistics. 19(2), 247-262. Leurgans, S. (1987). Linear models, random censoring and synthetic data, Biometrika, 74, 301-309. Breiman, L. (1996). Out-of-bag estimation, Technical report, Department of Statistics, University of California at Berkeley, CA, USA. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the RoyalStatistical Society. 58, 267-288

Sposto, R. (2002). Cure model analysis in cancer: an application to data from the Children's Cancer Group, Statistics in medicine. Volume 21, Issue 2, 293-312. Friedman, J., Hastie, T. and Tibshirani, R. (2007). Pathwise coordinate optimization, The Annals of Applied Statistics. 1, 302-332. Kleinbaum, D.G. and Klein, M. (2010). Survival Analysis, Springer. Hastie, T., Tibshirani, R. and Friedman, J. (2011). The Elements of Statistical Learning, 2nd Edition, Springer. Gail, M., Krickeberg, K., Samet, J.M., Tsiatis, A. and Wang, W. (2012). Survival Analysis, A Self-Learning Text, 3rd Edition, Springer. Zhou, Z.H. (2012). Ensemble Methods: Foundations and Algorithms, CRC Press, Boca Raton, FL. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2014). An Introduction to Statistical Learning with Applications in R, Springer. Nguyen, V. (2015). Mahalanobis kernel-based support vector data description for detection of large shifts in mean vector, Electronic Theses and Dissertations. 1160. Huh, M. (2015). Kernel-trick regression and classification. Communications for Statistical Applications and Methods. 22, 201-207.

Schölkopf, B. and Smola, A. J. (2002). Learning with Kernels, MIT Press. McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, 2nd Edition, Chapman & Hall. Freund, Y., Schapire, R. and Abe, N. (1999). A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence. 14(5), 771-780. Lee, S., Han, S. and Hwang, S. (2016). Ensemble approach for improving prediction in kernel regression and classification. Communications for Statistical Applications and Methods. 23, 355-362. Minh, H.Q., Niyogi, P. and Yao, Y. (2006). Mercer's theorem, feature maps, and smoothing. International Conference on Computational Learning Theory, COLT 2006: Learning Theory pp 154-168 Karatzoglou, A., Meyer, D. and Hornik, K. (2006). Support vector machines in R, Journal of statistical software Souza, C.R. (2010). Kernel functions for machine learning applications, Creative Commons Attribution-Noncommercial-ShareAlike 3.0, crsouza.com Sabin, C. and Petrie, A. (2019). Medical statistics at a glance, John Wiley & Sons, Ltd.

Chen, D.G.D., Peace, K.E. and Zhang, P. (2017). Clinical trial data analysis using R and SAS, Chapman & Hall Hosmer, D.W., Lemeshow, S., and May, S. (2008). Applied survival analysis: regression modeling of time-to-event data, Wiley-Interscience, New Jersey. Therneau, T. and Grambsch, P. (2000), Modeling Survival Data: Extending the Cox Model, Springer-Verlag, NewYork. Klein and Moeschberger (1997), Survival Analysis Techniques for Censored and truncated data, Springer. National Longitudinal Survey of Youth Handbook The Ohio State University, 1995. 한선우. (2016). 커널 능형 회귀분석에서 앙상블 기법을 이용한 효율성 연구 (석사학위 논문, 한국외국어대학교 대학원) 황성윤. (2017). 앙상블 기법을 이용한 커널 능형 로지스틱 회귀분류법의 효율성에 관한 연구 (석사학위 논문, 한국외국어대학교 대학원) 김종영. (2018). LASSO 벌점함수를 이용한로 버스트 회귀분석에서 회귀계수 추정에 대한 비교연구 (석사학위 논문, 한국외국어대학교 대학원)

김준영. (2018). 중도절단자료에서 유도된 비모수 회귀모형을 통한 분산 축소 (석사학위 논문, 한국외국어대학교 대학원) 이성희, (2018). 중도절단회귀모형에서의 변수선택법 (석사학위 논문, 한국외국어대학교 대학원) 이재병, (2020). 최신 부스팅 기법에 대한 연구 (석사학위 논문, 건국대학교 대학원) 국문초록 본 연구는 중도절단(censoring)이 포함된데이터에 대하여 회귀분석을 실시하는 경우 예측력(predictive power)을 향상시킬 수 있는 방법에 관한 것이다. 중도절단은 보통의학 분야에서 자주 등장하는 환자의 생존시간(survival time)과 관련한 생존자료(survival data)에서 환자가 연구대상인 질병 이외의 요인에 의해 사망하게 되는 등의 내부적 또는 외부적인 원인에 의하여 발생하게 된다. 생존자료를 분석하는가장 큰 목적은 어떠한 요인이 환자의 생존 시간에 유의미한 영향력을 미치게 되는지 확인하고 이를 통해 환자의 생존 시간을 예측하는 것이다. 이러한 중도절단이 포함된 생존자료의 경우는 추정의 대상이 되는 생존시간이 부분적으로만 관측되기 때문에이를 대체하기 위한 인조변수(synthetic response)를 만들어서 자료를 분석 할 수 있다.

하지만, 이러한 인조변수는 설명변수(explanatory variable)가 주어졌을 경우의 조건부 분산(conditional variance)이 원래 생존시간의 조건부분산보다 커지는 경향이 있고 생존시간이 증가할수록 증가하는 폭도 커지는 특성이 있다. 이 때문에 추정량에 대한 안정성이 떨어져서 문제가 될 수 있다. 이러한 문제점을 보완하기 위해 본 연구에서는 인조변수에 대한 회귀모형을 구축할 경우 변환함수를 따로 지정할 필요 없이 복잡한 비선형 데이터에 대해 적절한 사상함수를 사용해 설명변수 공간에 있는 데이터를 고차원의 특성 공간으로 이동시키는 커널트릭 기법(kernel trick method)과 다중공선성(multicollinearity)의 문제가 있을 때 적용 가능한 능형 회귀분석(ridge regression) 방법을 적용한다. 여기에 추가로 배깅(bagging) 및 랜덤포레스트(random forest)와 같은 앙상블 기법(ensemble method)을 적용하여 추정량의 분산을 줄임으로써 생존시간에 대한 예측력을 향상시키는 방법에 관하여 제안하고자 한다. 컴퓨터 모의실험을 통하여 다양한 상황을 가정하고 중도절단이 포함된 데이터에서 설명변수에 대한 예측력을 비교, 분석하였다. 이를 통해 본 연구에서 제안하고자 하는 방법이 일반적인 방법과 비교했을 때 전체적으로 우수한 예측력을 보임을 확인할수 있었다. 본 연구에서 제안하는 방법이다양한 연구 분야에서 나오는 중도절단이 포함된 데이터를 분석할 때 효율적으로 사용될 수 있기를 기대한다. 주요용어 : 생존자료, 생존분석, 생존시간, 중도절단, 인조변수, 능형 회귀분석, 기계학습, 커널트릭 기법, 앙상블 기법

문장표절률: **0%**

사사(謝辭)