

13강. 분류분석(1)

◆ 담당교수 : 이윤동 교수

■ 주요용어

용어	해설
선형판별분석	선형적인 판별식에 따라 판별을 수행하는 방법. 다변량 정규분포 가정에서, 분산공분산행렬의 동일성을 가정하는 경우에 얻어짐.
이차판별분석	다변량 정규분포 가정에서 얻어지는, 2차식 형태의 판별식을 이용한 판별 분석
신경망 방법	뉴런이 자극수용체의 자극을 전달받아 신호를 전달하 는 과정을 네트워크 형태의 정보처리과정으로 보고, 이를 모사하여 개발한 통계적 방법.
거칠기 벌칙	일관성 있는 예측 값의 생성을 위하여, 모형이 단순 화 되고 모수 개수가 줄어들도록 가한 벌칙항

■ 연습문제

1. 판별분석에서 독립변수에 대한 다변량 정규분포 가정을, 종속변수의 분포로 바꾸는데 이용하는 수학적 법칙은 무엇인가?

정답 및 해설 : 베이즈 정리

2. 로지스틱 회귀분석 방법에 대하여 선형판별분석 방법이 갖는 제약점은 무엇인가?

정답 및 해설 : 독립변수가 정규분포 가정을 할 수 있는 양적변수여야 한다.

3. 신경망 분류방법에서 히든노드의 개수가 계속 증가되면 어떤 현상이 일어나고 이런 현상을 제어하기 위한 대안은 무엇인가?

정답 및 해설 : 과적합문제가 발생하고, 거칠기벌칙항에 대한 가중치를 높여주는 방법으로 과적합 효과를 줄일 수 있다.

■ 참고자료

- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of ideas immanent in neural activity.
Bulletin of Mathematical Biophysics 5, 115-133.
- Rao, R. C. (1948). The utilization of multiple measurements in problems of biological classification.
Journal of the Royal Statistical Society, Series B 10 (2): 159-203.
- Ripley, B. D. (1993) Statistical aspects of neural networks.
In Networks and Chaos – Statistical and Probabilistic Aspects, eds O. E. Barndorff-Nielsen, J. L. Jensen and W. S. Kendall, pp. 40-123. London: Chapman & Hall.
- Ripley, B. D. (1994) Neural networks and flexible regression and discrimination.
In Statistics and Images 2, ed. K. V. Mardia, volume 2 of Advances in Applied Statistics, pp. 39-57. Abingdon: Carfax.
- Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S. Springer Science & Business Media.

■ 코스웨어

- R의 기본 코스웨어 학습방법 안내

다음은 R에 대한 기본적인 학습을 할 수 있는 방법을 안내 해 드립니다. 아직 R에 대하여 익숙하지 못한 학습자는 “R의 입문 코스웨어”를 통하여 R에 대하여 익숙해지기 바랍니다.

1) [학교 홈페이지에 접속] - [우측하단 “학과” 선택]



2) ["정보통계학과" 선택]



3) [상단메뉴 학습정보 내의 “학습자료실” 선택]



4) [좌측메뉴의 “코스웨어” 선택]



5) [좌측메뉴의 “R의 입문 코스웨어” 선택]



6) “R의 입문 코스웨어”는 총10장으로 구성됨



7) 각 장의 좌측메뉴의 학습목차를 참고하여 학습



“R의 입문 코스웨어” 과정을 통하여 R의 기본개념 및 구조를 이해하고 실제 작업에 필요한 자료처리 및 분석기법을 응용할 수 있는 능력을 학습하기 바랍니다.