

1. "Bankruptcy.csv"파일은 총 46개의 현재 재무구조가 건실한 기업과 파산한 기업들에 대한 2년전 재무상황자료이다. 이 자료를 이용하여 건실기업과 파산기업을 판별하시오.

X1 : Cash Flow/Total Debt	X2 : Net Income/Total Asset
X3 : Current Asset/Current Liabilities	X4 : Current Asset/Net Sales
group : 1=파산기업 , 2=건실기업	

- (1) 두 집단의 공분산이 같다고 할 수 있는지 유의수준 5%에서 가설검정 하시오.

가설

$$H_0 : \Sigma_1 = \Sigma_2 \quad H_1 : \Sigma_1 \neq \Sigma_2$$

검정결과

χ^2 -통계량=64.049>18.307= $\chi^2_{0.05}(10)$ (p-값=6.159e-10<0.05= α)이므로 귀무가설을 기각
즉, 두 집단의 공분산행렬은 동일하지 않다.

- (2) 위 (1) 검정결과를 이용하여 선형판별함수 또는 이차판별함수 중 선택하여 추정하시오(변수선택법 사용, criterion="AC"). 판별결과를 이용하여 분류표 작성과 정분류율을 계산하시오.

분류표

관측 \ 예측	파산기업	건실기업
파산기업	19	1
건실기업	2	24

정분류율
약 93.5%

- (3) 로지스틱 회귀모형식을 작성하시오.(단계적 변수선택법 사용)

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = -5.940 + 6.556X_1 + 3.019X_3$$

- (4) 추정된 로지스틱 회귀모형식에서 각 독립변수의 오즈 증가량을 구하고 이에 대한 해석을 작성하시오.

절편	X_1	X_3
0.003	703.744	20.473

X_1 이 1단위 증가할 때 마다 파산기업에 대한 건설기업의 비인 오즈가 약 703배 증가
즉, X_1 이 높으면 건설기업일 경향이 높음

X_3 이 1단위 증가할 때 마다 파산기업에 대한 건설기업의 비인 오즈가 약 20배 증가
즉, X_3 이 높으면 건설기업일 경향이 높음

- (5) 로지스틱 회귀모형식에 의해서 판별한 결과를 이용하여 분류표 작성과 정분류율을 계산하시오.

분류표

관측 \ 예측	파산기업	건설기업
파산기업	18	1
건설기업	3	24

정분류율
약 91.3%

2. “voter.txt”는 미국 대선 후보의 지지자들에 대한 조사 자료이다. 이 자료를 이용하여 연령과 성별, 학력수준에 따라서 어떤 후보를 지지하는지 알아보고자 한다. 전체 자료 중에서 학습자료 60%와 검정자료 40%로 나누어서 작업하시오.(단, 학습자료, 검정자료의 표본추출시 `set.seed(12345)`)

candidate : 후보	educ : 피교육년
age : 연령	gender : 성별

- (1) Naive Bayes 분류방법을 이용하여 분류해보시오.

검정자료 분류표

관측 \ 예측	버니샌더스	트럼프	힐러리
버니샌더스	7	15	8
트럼프	41	89	104
힐러리	56	149	269

검정자료 정분류율
약 49.5%

(2) 연령(age)과 피교육년수(educ)를 그룹화하여 위와 동일하게 Naive Bayes 분류방법을 이용하여 분류해보시오.

연령	연령대	피교육년	학력
age<30	30대 이하	educ<10	고등학교 미만
30≤age<40	30대	10≤educ<13	고등학교
40≤age<50	40대	13≤educ<16	전문대학교
50≤age<60	50대	16≤educ<18	대학교
age≥60	60대 이상	educ≥18	대학원

검정자료 분류표

관측 \ 예측	버니샌더스	트럼프	힐러리
버니샌더스	0	0	0
트럼프	36	72	81
힐러리	68	181	300

검정자료 정분류율
약 50.4%

3. “EgyptSkull.csv”은 이집트인의 두개골에 대한 자료이다. 이 자료를 이용하여 연대(Year)를 예측할 수 있는 최적의 SVM(모수조율 방법)과 신경망 모형(nnet 패키지)을 작성하시오. 전체 자료 중에서 학습자료 70%와 검정자료 30%로 나누어서 작업하시오.(단, 학습자료, 검정자료의 표본추출시와 신경망 훈련시 set.seed(12345))

No. : 고유번호
X1 : Maximum Breadth(mm) 머리뼈의 둘레 중 가장 큰 둘레
X2 : Basibregmatic Height(mm) 기저시상봉합과 관상봉합의 접합점의 크기로 머리뼈의 정수리부터 눈썹뼈 위까지의 길이
X3 : Basialveolar Length(mm) 기조치조의 길이로 맨 앞의 치아부터 혀가 닿는 끝부분까지의 길이
X4 : Nasal Height of Skull(mm) 코의 높이
Year : 두개골 형성 시기의 대략적인 연도 기원전 4000년, 기원후 150년

SVM 모수 조율	신경망 모수 설정
kernel : radial	hidden node size : 5
cost : seq(100, 1000, length=21)	rang : 0.00001
gamma : seq(0.000001, 0.1, length=31)	decay : 0.00001
	maxit : 500

검정자료 SVM 분류표

관측 예측	-4000	-3300	-1850	-200	150
-4000	2	5	1	1	1
-3300	3	2	1	5	2
-1850	1	3	2	0	0
-200	1	0	2	3	4
150	1	0	1	3	1

검정자료 신경망 분류표

관측 예측	-4000	-3300	-1850	-200	150
-4000	4	8	0	3	1
-3300	0	0	0	0	0
-1850	3	2	5	4	2
-200	0	0	0	0	0
150	8	10	7	12	8

