판별분석 (분류분석)

Discriminant Analysis (Classification Analysis)

판별분석이란?

주로 logistic Regression.

- 분류되어 있는 집단 간의 차이를 의미있게 설명해 줄 수 있는 독립변수들을 찾음.
- 변수의 결합으로 판별식(Discriminant function)을 만들어 관측치를 판별하는 규칙 만듬.
- 예
  - 서비스 이용 불만 고객의 성향 분석 → 종속변수의 유무에 따라
  - SKT/KT/LGT 가입고객 판별 변수 및 판별함수 유도
- 비교: 군집분석
  - 관측치를 그룹으로 구분함
  - 판별분석과는 다르게 데이터에 집단을 나타내는 변수 없음 즉, 정답X.
- 비교: 회귀분석
  - 종속변수가 이산형인 logistic regression은 두 집단 판별분석의 한 방법
  - 판별분석은 집단이 2개 이상의 범주를 갖는 범주형변수인 경우 가능
  - 판별분석은 집단을 구분하는 판별식 유도
  - 회귀분석은 집단에 속하는 확률 예측

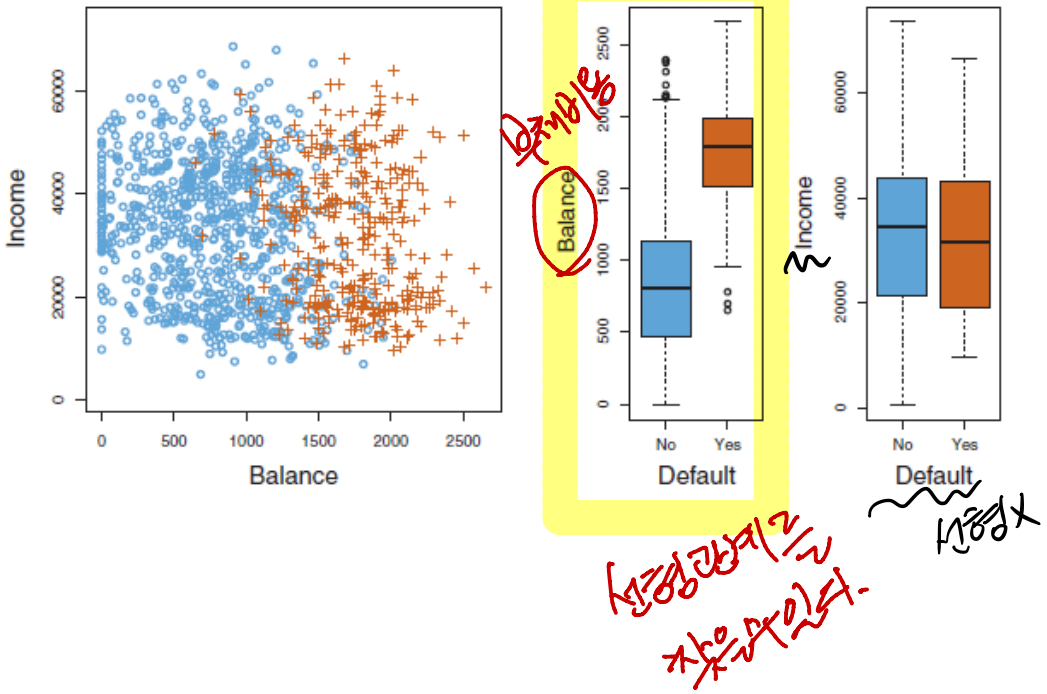
Clustering vs. classification.

Binary y ⇒ logistic-

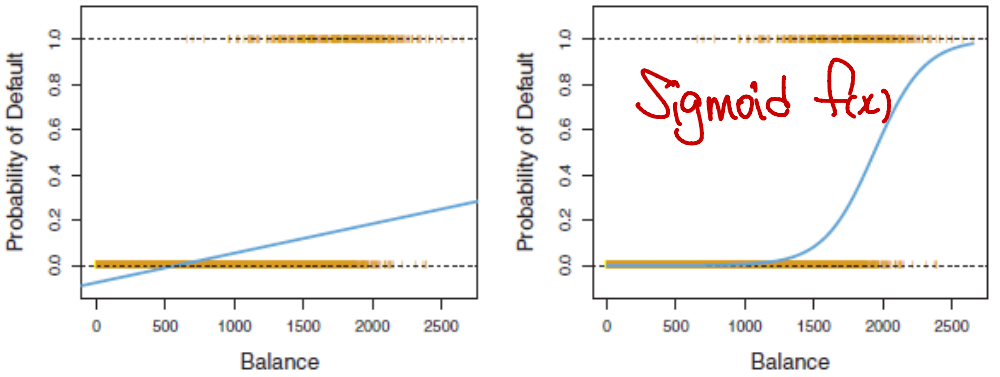
↳ multi normal classification

Default Dataset

- 개인의 연봉과 월 신용카드 잔고를 사용해 파산 예측



- 선형회귀모형은 적당치 않음
  - 예측치가 [0,1] 밖에 있을 수 있음



Simple Logistic Regression

- Balance를 사용해 default=Yes일 확률 예측

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

$$\Leftrightarrow \log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

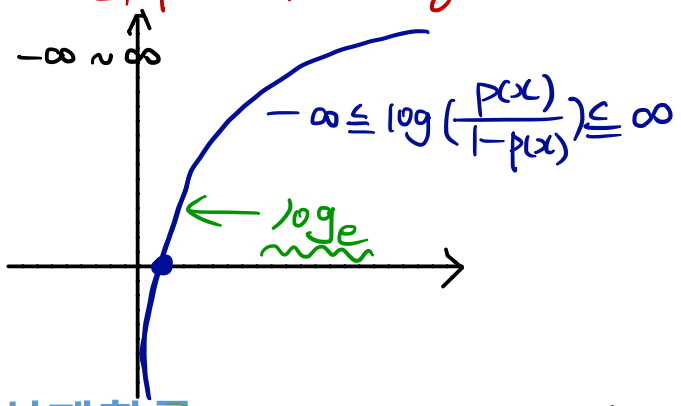
log-odds (logit)

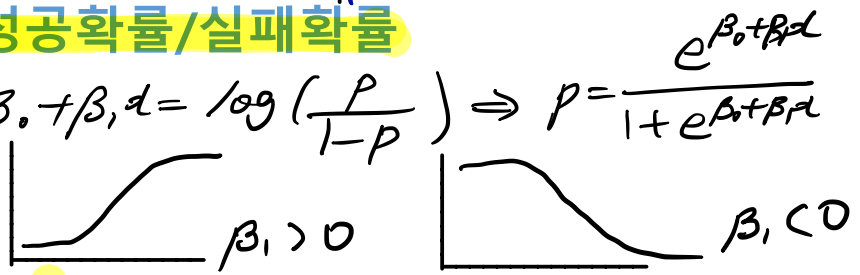
Odds: 성공확률/실패확률

- X가 한단위 증가할 때 log-odds가 β_1 만큼 증가

➔ odds가 e^{β_1}배 변화

- $p(X)$와 X가 선형관계가 아니기 때문에 X의 한단위 증가가 일으키는 p(X)의 증가는 X의 값에 따라 달라짐
- $\beta_1 > 0$: X가 증가하면 p(X)가 증가
- $\beta_1 < 0$: X가 증가하면 p(X)가 감소





$y = \beta_0 + \beta_1 x + \epsilon$

if $x=1$ 여자 $\Rightarrow$ 여 : $\beta_0 + \beta_1 + \epsilon$

0 남자 남 : $\beta_0 + \epsilon$

Linear.

여 $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1$

남 $\log\left(\frac{p}{1-p}\right) = \beta_0$

e를 씌우고

$\Rightarrow \frac{p}{1-p} = e^{\beta_0+\beta_1}$

$\frac{p}{1-p} = e^{\beta_0}$

$\frac{\text{odds여}}{\text{odds남}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = \frac{e^{\beta_0} \times e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$

odds비

↳ 여기서는 0, 1 이지만 (남 여)

Continuous한 수치인경우.

1단위↑ $\Rightarrow e^{\beta_1}$ 배증가

odds가.

- 회귀 계수 추정: Maximum likelihood Method

$$L(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} (1 - p(x_{i'}))$$

```
> library(ISLR)
> glm.fit=glm(default~balance,data=Default,family=binomial)
> summary(glm.fit)

Call:
glm(formula = default ~ balance, family = binomial, data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2697  -0.1465  -0.0589  -0.0221   3.7589

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.065e+01  3.612e-01  -29.49   <2e-16 ***
balance      5.499e-03  2.204e-04   24.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1596.5  on 9998  degrees of freedom
AIC: 1600.5

Number of Fisher Scoring iterations: 8
```

$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -10.65 + 0.005 \times \text{balance}$.

- balance의 회귀계수 p-value<0.0001
  - 파산확률과 카드잔고 사이에 관계가 있음
- 카드 잔고가 $1000인 사람의 파산 확률은?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 1{,}000}}{1+e^{-10.6513+0.0055\times 1{,}000}} = 0.00576,$$

Value (1000원) / log(P) = ⇒ 로그

```
> a=predict(glm.fit,data.frame(balance=1000))
> exp(a)/(1+exp(a))
          1
0.005752145
> predict(glm.fit,data.frame(balance=1000),type="response")
          1
0.005752145
```

파산확률 0.06%.

Multiple Logistic Regression

각 항목의 독립변수가 고정이라고 할때.
즉 x_i가 1일때이다. β_i의 영향도.

- 모형

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p,$$

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}.$$

```
> glm.fit2=glm(default~balance+income+student,data=Default,family=binomial)
> summary(glm.fit2)

Call:
glm(formula = default ~ balance + income + student, family = binomial,
    data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4691  -0.1418  -0.0557  -0.0203   3.7383

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.087e+01  4.923e-01 -22.080  < 2e-16 ***
balance      5.737e-03  2.319e-04  24.738  < 2e-16 ***
income       3.033e-06  8.203e-06   0.370  0.71152
studentYes  -6.468e-01  2.363e-01  -2.738  0.00619 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1571.5  on 9996  degrees of freedom
AIC: 1579.5

Number of Fisher Scoring iterations: 8
```

- 현재잔고와 학생여부는 파산과 연관이 있음
- 학생더미 변수의 계수가 음수
  - Simple logistic regression에서는 양수

```
> glm.fit3=glm(default~student,data=Default,family=binomial)
> summary(glm.fit3)

Call:
glm(formula = default ~ student, family = binomial, data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2970  -0.2970  -0.2434  -0.2434   2.6585

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.50413    0.07071  -49.55  < 2e-16 ***
studentYes   0.40489    0.11502    3.52 0.000431 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

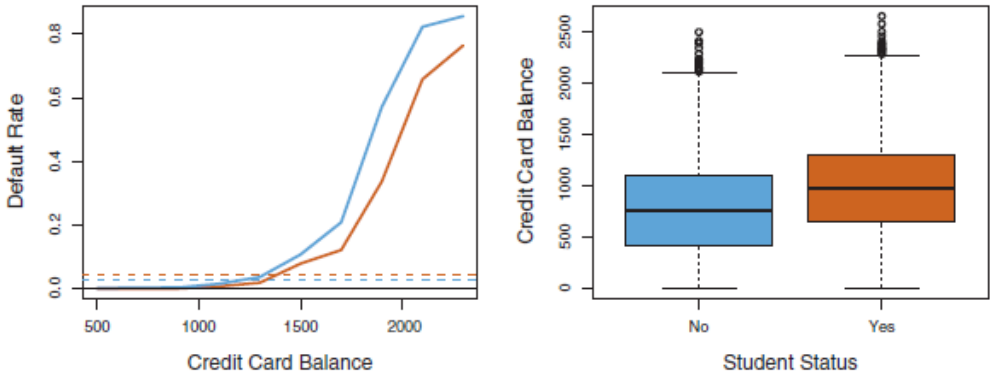
- 학생이면 파산가능성이 더 낮음?

- Confounding effect
  - 같은 수준의 income과 balance를 가지고 있을 경우 학생의 파산가능성(odds)이 더 낮음
  - 학생일 경우 credit card balance가 더 높은 경향이 있음

  - 학생인 경우 아닌 경우보다 파산가능성 (odds) 이 1.5배
  - 같은 수준의 income과 balance를 가지고 있을 경우 학생의 파산가능성 (odds) 은 0.5배

```
> exp(coef(glm.fit2)[4])
studentYes
 0.5237317
```

```
> exp(coef(glm.fit3)[2])
studentYes
  1.499133
```



모형비교: Deviance Goodness-of-fit Test

```
> anova(glm.fit2, glm.fit3, test="Chisq")
Analysis of Deviance Table

Model 1: default ~ balance + income + student
Model 2: default ~ student
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      9996     1571.5
2      9998     2908.7 -2  -1337.1 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Reduced Model과 Full Model의 차이가 유의한지 검정
- 여러 설명변수가 주는 영향이 유의한지 한번에 검정
  - 다중회귀분석의 F-test와 유사
- Balance와 income의 영향이 default 확률을 예측하는 데 유의하다.

반복 측정된 자료

- 한 개의 X값에서 여러 개의 Y가 측정된 경우
- X_i에서의 관측치가 0, 1이 아니라 n_i개 중 $Y_{\cdot i}$개의 성공 관측
- Binomial Distribution

$$f(Y_{\cdot i}) = \frac{n_i!}{Y_{\cdot i}!\,(n_i - Y_{\cdot i})!}\pi_i^{Y_{\cdot i}}(1-\pi_i)^{n_i - Y_{\cdot i}}$$

Example: Coupon Effectiveness

- 가격을 할인해 주는 쿠폰의 효과를 검증하기 위해 무작위로 추출된 각 200개의 가구에 5,10,15,20,30 달러의 쿠폰을 제공했다.

	Price_reduc	N	N_redeemed
1	5	200	30
2	10	200	55
3	15	200	70
4	20	200	100
5	30	200	137

```
> data=read.csv("coupon.csv")
> model2=glm(cbind(N_redeemed,N-N_redeemed)~Price_reduc,data=data,family=binomial(logit))
> summary(model2)

Call:
glm(formula = cbind(N_redeemed, N - N_redeemed) ~ Price_reduc,
    family = binomial(logit), data = data)

Deviance Residuals:
      1       2       3       4       5
-0.8988  0.6677 -0.1837  0.7612 -0.5477

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.044348   0.160977  -12.70   <2e-16 ***
Price_reduc  0.096834   0.008549   11.33   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 149.4627  on 4  degrees of freedom
Residual deviance:   2.1668  on 3  degrees of freedom
AIC: 33.793

Number of Fisher Scoring iterations: 3
```

$$\hat{\pi} = \frac{\exp(-2.04 + 0.0968X)}{1 + \exp(-2.04 + 0.0968X)}$$

$\widehat{OR} = \exp(0.0968) = 1.102$➔ 쿠폰의 할인액이 1달러 증가할 때 쿠폰을 사용할 Odds가 10% 증가한다.

오분류율, 민감도, 특이도

- 판별분석 결과 분할표

		예측그룹	
실제그룹	표본의 수	Positive	Negative
Positive	n_1	n_{11}	n_{12}
Negative	n_2	n_{21}	n_{22}

- 오분류율 (Error rate): 잘못 분류한 비율 $= \frac{n_{12}+n_{21}}{n_1+n_2}$
- 민감도 (Sensitivity: True Positive Rate):positive를 positive로 구분한 비율 $= \frac{n_{11}}{n_1}$
- 특이도 (Specificity: True Negative Rate): negative를 negative로 구분한 비율$= \frac{n_{22}}{n_2}$

Example: 오분류율, 민감도, 특이도

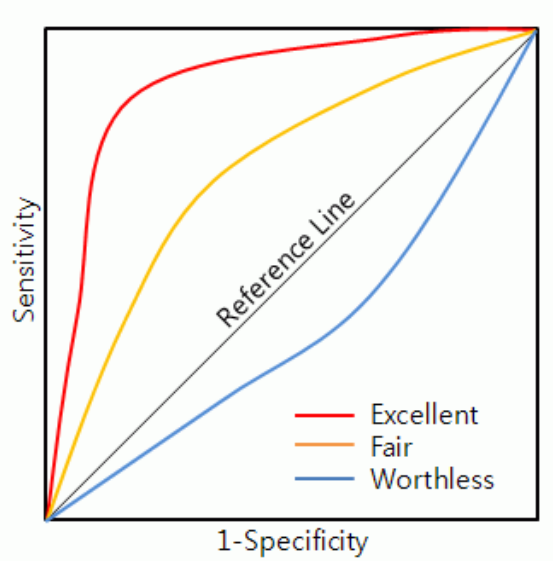
```
> pred=data.frame(default=Default$default, fit=glm.fit2$fitted)
> head(pred)
  default          fit
1      No 0.0014287239
2      No 0.0011222039
3      No 0.0098122716
4      No 0.0004415893
5      No 0.0019355062
6      No 0.0019895182

> table(Default$default)

  No  Yes
9667  333
> xtabs(~Default$default+(glm.fit2$fitted>0.5))
              glm.fit2$fitted > 0.5
Default$default FALSE TRUE
            No   9627   40
            Yes   228  105
```

- 10000 명중 330명이 파산
- Cutoff=0.5

- Sensitivity= 105/(228+105)=0.315
- Specificity=9627/(9627+40)=0.996
- Error rate=(40+228)/10000=0.0268

ROC (Receiver Operating Characteristic) Curve

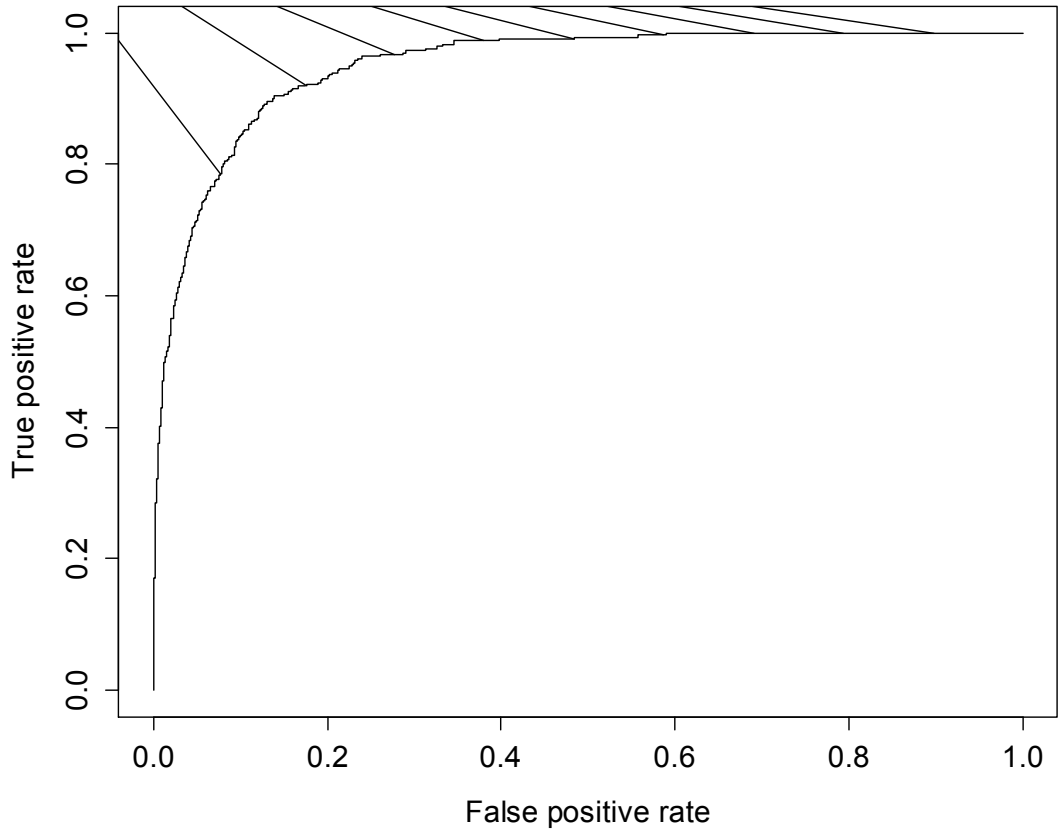
- y축: 민감도 =True positive rate
- x축: 1- 특이도=False positive rate



- ROC curve의 아래쪽 면적(AUC)이 클수록 좋은 모형
- 왼쪽 코너에 가까운 포인트를 Cutoff로 정하는 것도 한 방법

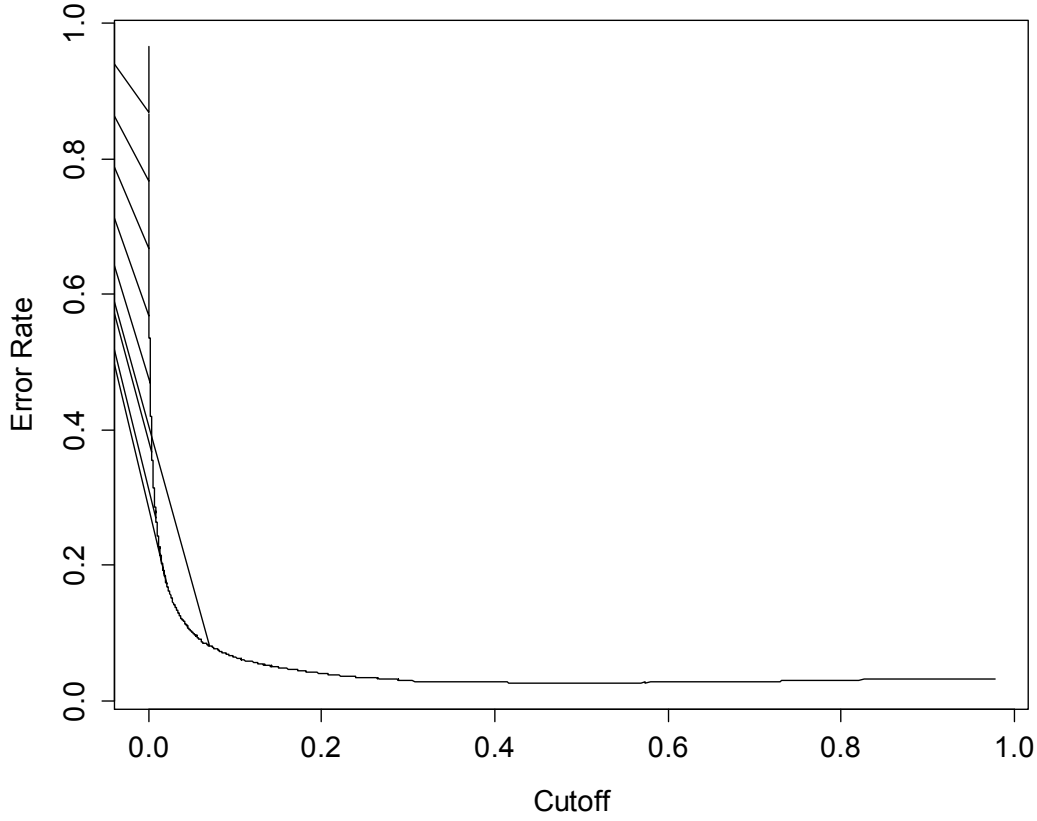
Example: ROC curve for Default Dataset

```
library(ROCR)

predob=prediction(pred$fit,pred$default)
plot(performance(predob,"tpr","fpr"))
```



Example: Error rate for Default Dataset

```
plot(performance(predob,measure="err"))
```



Linear Discriminant Analysis (LDA)

- π_k: k-집단에 들어갈 사전 확률(prior probability). $\sum_k \pi_k = 1$
- 데이터가 주어져 있을 때 k-집단에 들어갈 사후 확률 (posterior probability)

$$\Pr(G = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}$$

- 각 집단의 분포가 정규분포라고 가정

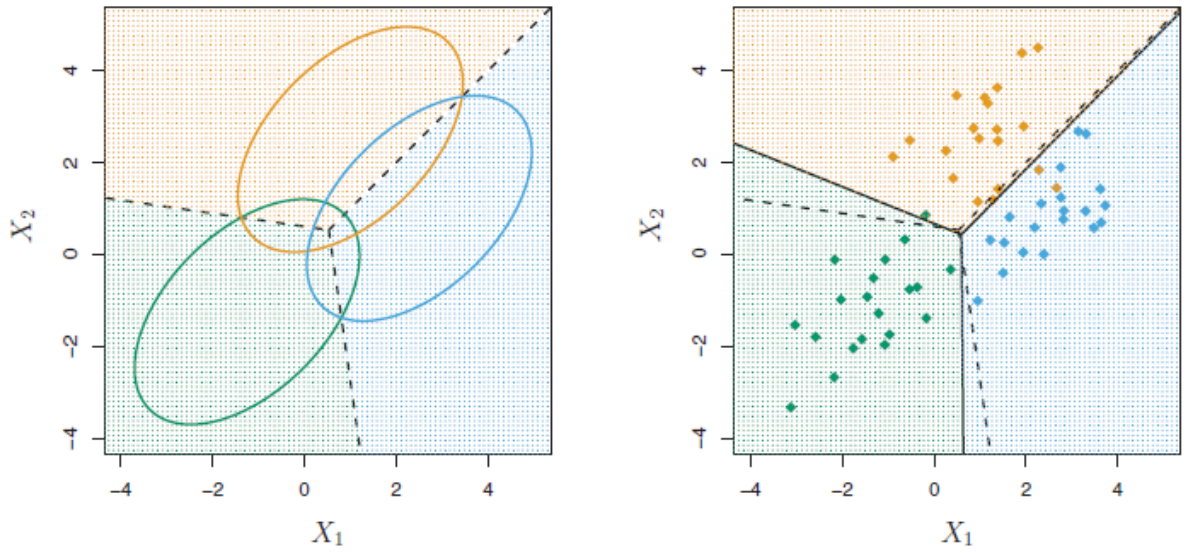
$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

- 만약 각 집단의 공분산 행렬이 동일하다면 ($\Sigma_k = \Sigma$), 사후확률의 비율 $\log\frac{\Pr(G=k|X=x)}{\Pr(G=l|X=x)}$이 x 에 대한 선형식으로 표현됨.
- 판별함수 (discriminant function)이 큰 그룹으로 할당

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k$$

$$x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k = x^T\Sigma^{-1}\mu_l - \frac{1}{2}\mu_l^T\Sigma^{-1}\mu_l$$

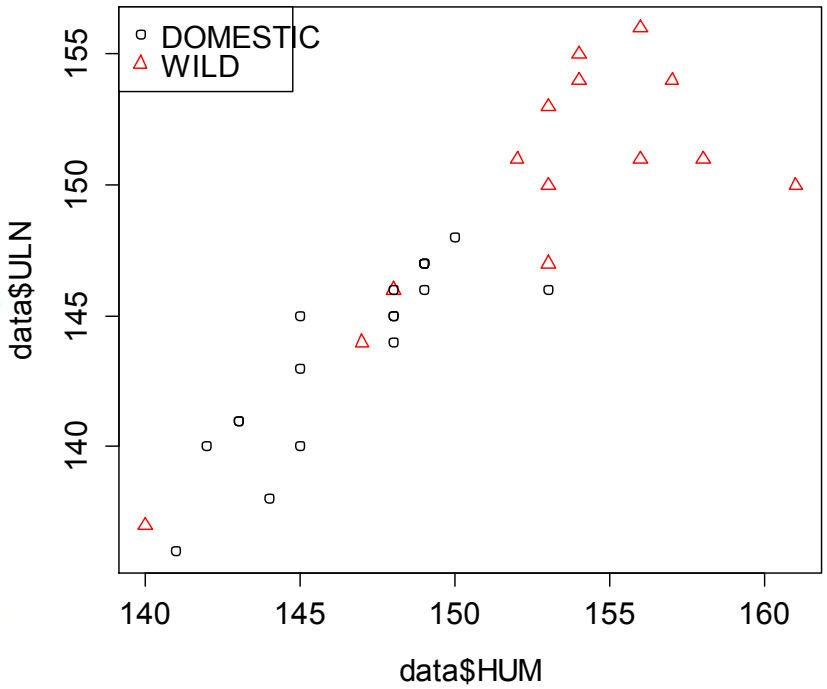
- 위의 식을 만족하는 선이 decision boundary



예: Turkey

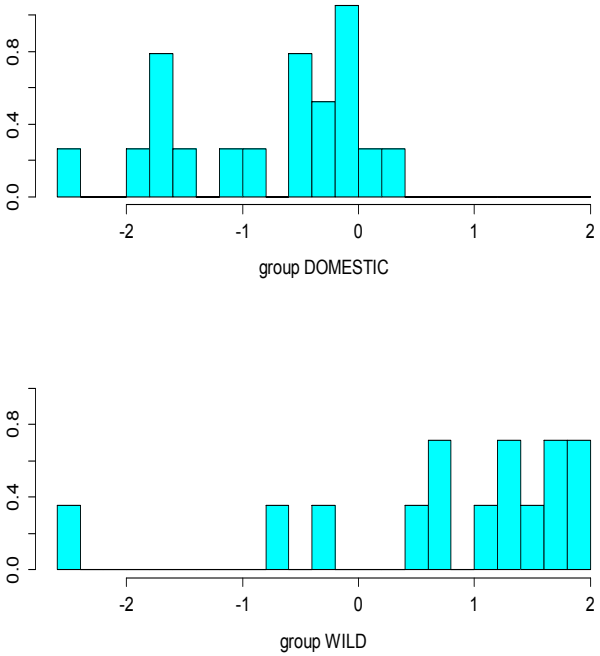
- 야생칠면조(WILD)와 사육칠면조(DOMESTIC)의 구분
- 9개 부위의 길이 정보를 사용

```
> head(data)
     ID HUM RAD ULN FEMUR TIN CAR D3P COR SCA TYPE
13 B710 153 140 147   142 151 817 305 102 128 WILD
14 B790 156 137 151   146 155 814 305 111 137 WILD
17 B819 158 135 151   146 152 790 289 111 125 WILD
19 B085 148 129 146   139 147 767 287 106 123 WILD
20 B089 157 140 154   140 159 818 301 116 136 WILD
21 B090 153 138 153   141 151 822 312 115 133 WILD
> str(data)
'data.frame': 33 obs. of  11 variables:
 $ ID   : Factor w/ 82 levels "B081","B085",..: 20 21 24 2 3 4 5 7 8
 $ HUM  : int  153 156 158 148 157 153 156 153 152 147 ...
 $ RAD  : int  140 137 135 129 140 138 138 135 140 130 ...
 $ ULN  : int  147 151 151 146 154 153 156 150 151 144 ...
 $ FEMUR: int  142 146 146 139 140 141 145 144 144 136 ...
 $ TIN  : int  151 155 152 147 159 151 150 158 158 145 ...
 $ CAR  : int  817 814 790 767 818 822 835 772 792 765 ...
 $ D3P  : int  305 305 289 287 301 312 310 276 303 289 ...
 $ COR  : int  102 111 111 106 116 115 118 102 111 108 ...
 $ SCA  : int  128 137 125 123 136 133 133 123 122 131 ...
 $ TYPE : Factor w/ 2 levels "DOMESTIC","WILD": 2 2 2 2 2 2 2 2 2 2
 - attr(*, "na.action")=Class 'omit'  Named int [1:49] 1 2 3 4 5 6 7
  .. ..- attr(*, "names")= chr [1:49] "1" "2" "3" "4" ...
```



예: Turkey

```
> model1=lda(TYPE~HUM+ULN,data)
> model1
Call:
lda(TYPE ~ HUM + ULN, data = data)

Prior probabilities of groups:
 DOMESTIC      WILD
0.5757576 0.4242424

Group means:
              HUM      ULN
DOMESTIC 146.7368 143.7895
WILD     153.0000 149.9286

Coefficients of linear discriminants:
          LD1
HUM 0.1469784
ULN 0.1028563
```



예: Turkey

- 새로운 관측치 분류

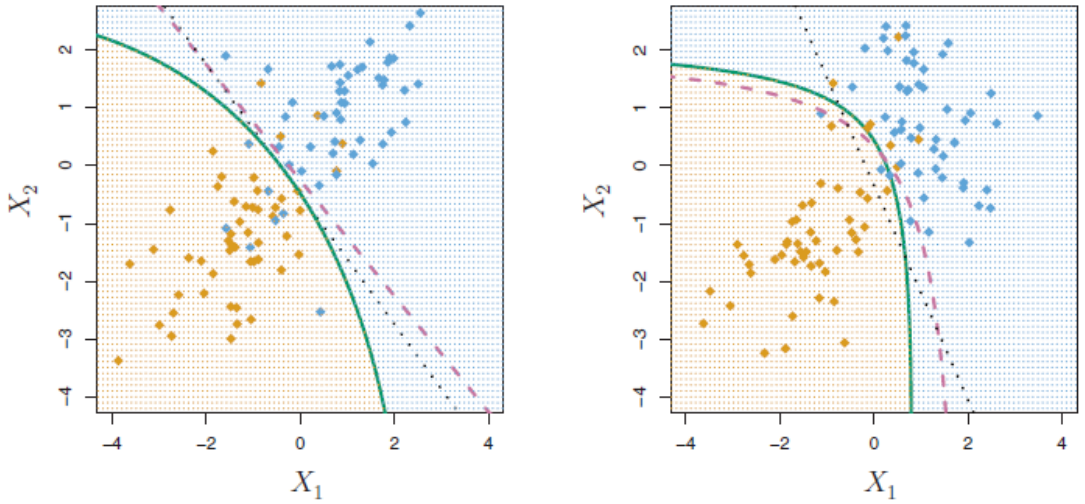
```
> predict(fit1,data.frame("HUM"=c(145,150),"ULN"=c(150,145)))
$class
[1] DOMESTIC DOMESTIC
Levels: DOMESTIC WILD

$posterior
   DOMESTIC      WILD
1 0.7139229 0.2860771
2 0.6392539 0.3607461

$x
          LD1
1 -0.27490816
2 -0.05429765
```

Quadratic Discriminant Analysis (QDA)

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)}$$

- 만약 각 집단의 공분산 행렬이 동일하지않다면 ($\Sigma_k \neq \Sigma$), 사후확률의 비율 $\log\frac{\Pr(G=k|X=x)}{\Pr(G=l|X=x)}$이 x 에 대한 이차식으로 표현됨.

$$\begin{aligned}\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^T\mathbf{\Sigma}_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\mathbf{\Sigma}_k| + \log\pi_k \\ &= -\frac{1}{2}x^T\mathbf{\Sigma}_k^{-1}x + x^T\mathbf{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\mu_k^T\mathbf{\Sigma}_k^{-1}\mu_k - \frac{1}{2}\log|\mathbf{\Sigma}_k| + \log\pi_k\end{aligned}$$



- 두 집단 공분산의 차이가 적고, p가 적당히 크며(p>6), 표본의 수가 작으면 (n_1, n_2 <25) QDF는 LDF 보다 비효율적

```
> fit2=qda(TYPE~HUM+ULN,data)
> fit2
Call:
qda(TYPE ~ HUM + ULN, data = data)

Prior probabilities of groups:
 DOMESTIC      WILD
0.5757576 0.4242424

Group means:
              HUM      ULN
DOMESTIC 146.7368 143.7895
WILD     153.0000 149.9286
> predict(fit2,data.frame("HUM"=c(145,150),"ULN"=c(150,145))) #prediction
$class
[1] WILD     DOMESTIC
Levels: DOMESTIC WILD

$posterior
    DOMESTIC      WILD
1 0.00221008 0.9977899
2 0.68425690 0.3157431
```

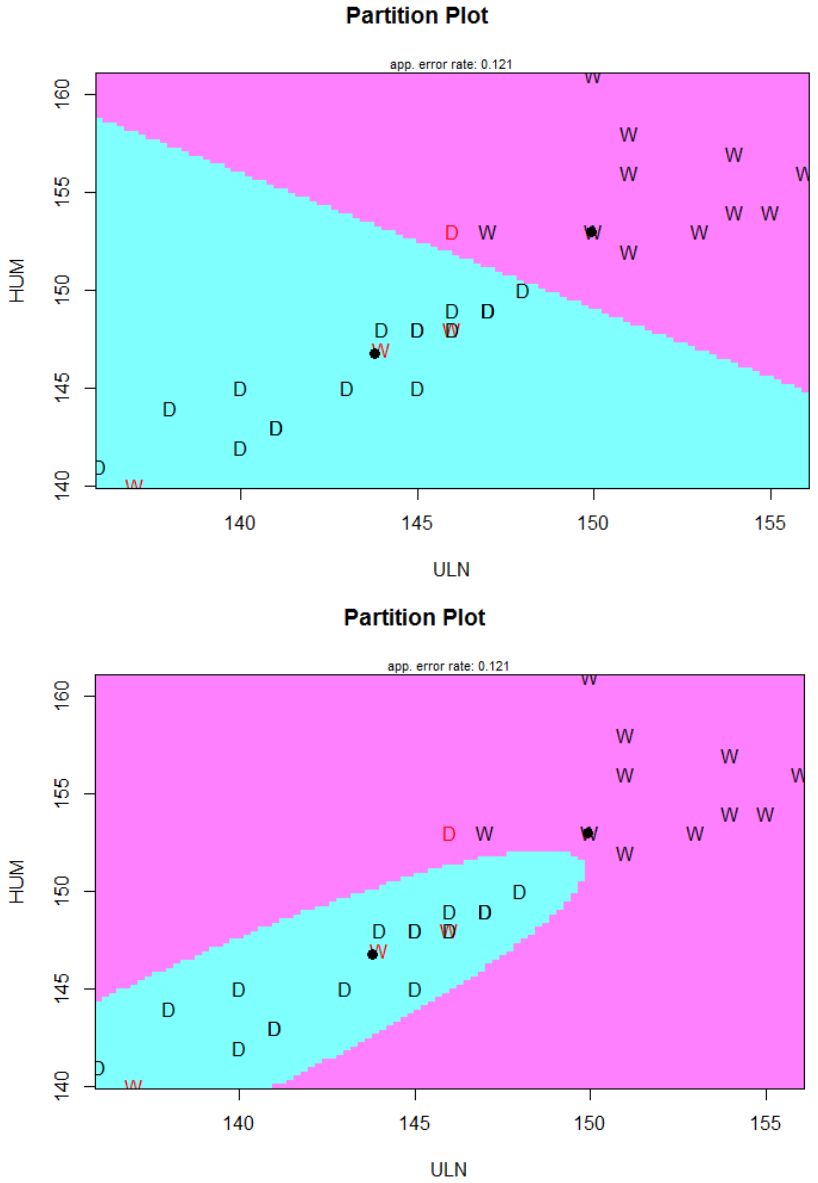
오분류율 계산

- 표본분할에 의한 오류율 계산
  - 동일한 데이터로 분류함수를 구하고 다시 오류율을 계산하면 실제보다 bias가 작게 계산
  - 자료를 training set과 validation set으로 나눔
  - training set: 판별함수 계산
  - validation set: 타당성 검사를 위한 표본으로 오류율 계산
- 교차타당성(cross-validation)에 의한 오류율 계산
  - 한 개만의 표본을 제외한 나머지 표본으로 판별함수 계산
  - 구해진 판별함수를 이용해 제외된 표본을 분류
  - 이 과정을 전체 표본 크기만큼 시행하여 오류율 계산

예: Turkey – 오분류율 계산

- CV=TRUE 옵션: 교차타당성에 의한 분류결과

```
> fit3=lda(TYPE~HUM+ULN,data,CV=T)
> ct3=table(data$TYPE,fit3$class)
> ct3

          DOMESTIC WILD
  DOMESTIC       18    1
  WILD            3   11
> prop.table(ct3)

             DOMESTIC       WILD
  DOMESTIC 0.54545455 0.03030303
  WILD     0.09090909 0.33333333
> sum(diag(prop.table(ct3)))
[1] 0.8787879


> library(klaR)
> partimat(TYPE~HUM+ULN,data,method="lda")
> partimat(TYPE~HUM+ULN,data,method="qda")
```



예: Turkey – 사전확률 설정

- 설정 하지 않으면 default로 자료가 가지는 각 그룹의 비율을 사전확률로 여김

```
> fit1
Call:
lda(TYPE ~ HUM + ULN, data = data)

Prior probabilities of groups:
 DOMESTIC      WILD
0.5757576 0.4242424

Group means:
              HUM      ULN
DOMESTIC 146.7368 143.7895
WILD     153.0000 149.9286

Coefficients of linear discriminants:
          LD1
HUM 0.1469784
ULN 0.1028563
```

- 사전 정보에 의해 Domestic일 확률이 0.4, Wild일 확률이 0.6인 것을 알 경우

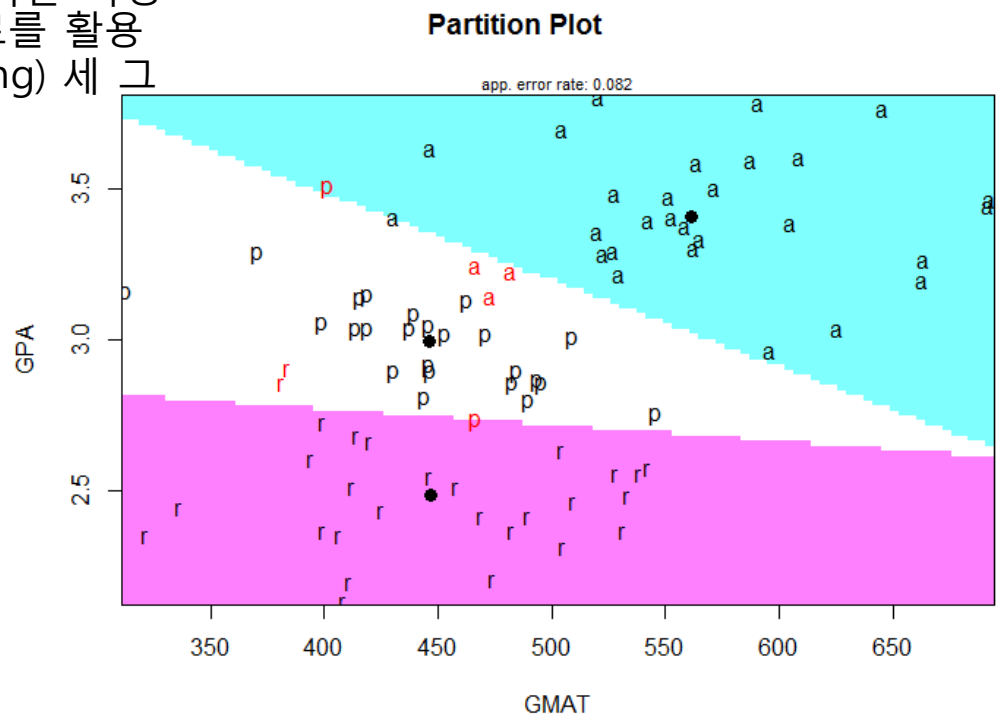
```
>
> fit4=lda(TYPE~HUM+ULN,data,CV=T,prior=c(0.4,0.6))
> ct4=table(data$TYPE,fit4$class)
> ct4

          DOMESTIC WILD
  DOMESTIC       12    7
  WILD            3   11
> sum(diag(prop.table(ct4)))
[1] 0.6969697
```

여러그룹의 판별분석

- 공분산 행렬이 그룹 간에 모두 같다는 가정➔ LDA
- 공분산 행렬이 그룹 간에 다르다는 가정 ➔ QDA
- 각 그룹에 속할 사후 확률이 가장 큰 집단으로 분류
- 예) 미국의 한 경영대학원에서 MBA과정에 지원하는 학생들의 학부 평균성적과 경영대학원 입학성적 자료를 활용해 합격(accepted), 불합격(rejected), 보류(pending) 세 그룹으로 분류

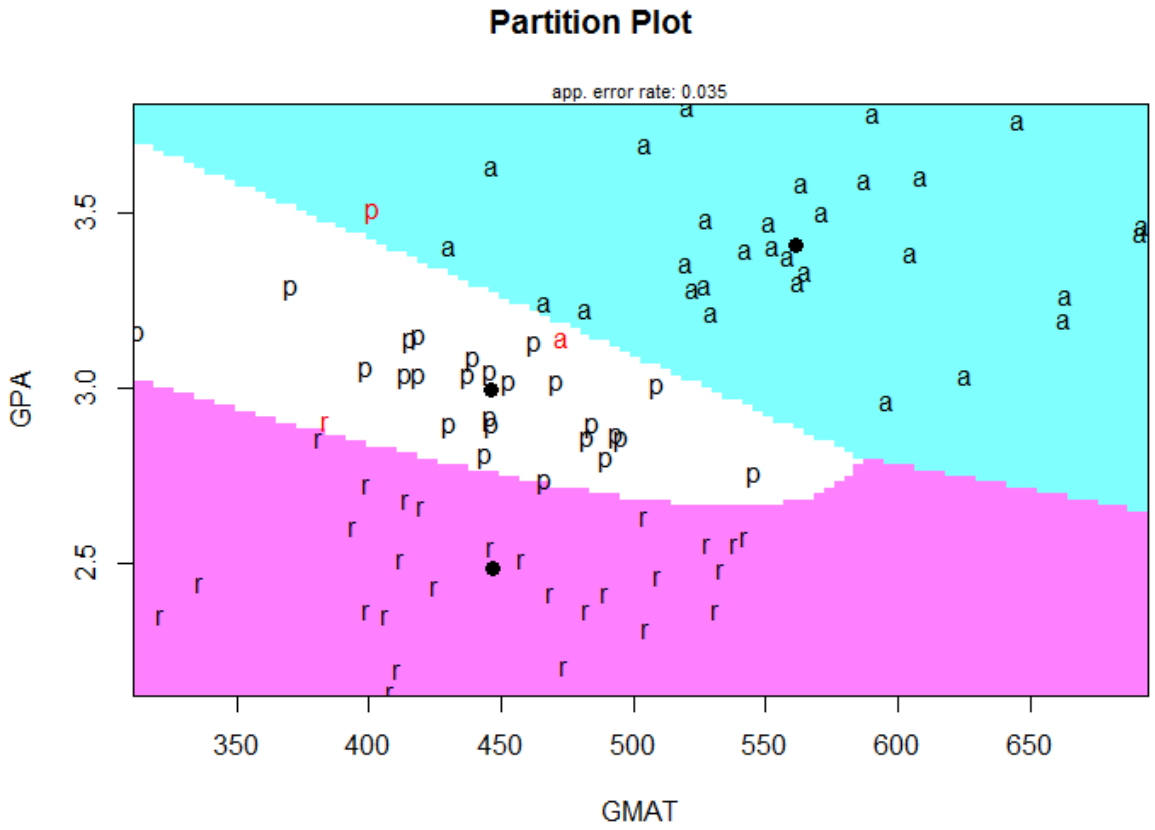
```
> head(mba)
   GPA GMAT   result
1 2.96  596 accepted
2 3.14  473 accepted
3 3.22  482 accepted
4 3.29  527 accepted
5 3.69  505 accepted
6 3.46  693 accepted
```

```
> fit5=lda(result~GPA+GMAT,mba,CV=T)
> ct5=table(mba$result,fit5$class)
> ct5

           accepted pending rejected
  accepted       27       4        0
  pending         1      24        1
  rejected        0       2       26
> sum(diag(prop.table(ct5)))
[1] 0.9058824
> partimat(result~GPA+GMAT,mba,method="lda")
```



여러그룹의 판별분석

```
> fit6=qda(result~GPA+GMAT,mba,CV=T)
> ct6=table(mba$result,fit6$class)
> ct6

           accepted pending rejected
  accepted       30       1        0
  pending         1      24        1
  rejected        0       1       27
> sum(diag(prop.table(ct6)))
[1] 0.9529412
> partimat(result~GPA+GMAT,mba,method="qda")
```



공분산 행렬의 동일성 검정

- Box's M Test
  - 가설
    - $H_0 : \Sigma_1 = \cdots = \Sigma_G = \Sigma$
    - $H_a : Not\ H_0$
  - 검정통계량

$$M = \prod \left(\frac{|C_{w(g)}|}{|C_w|} \right)^{(n_g - 1)/2}$$

    - $C_{w(g)}$: 그룹g의 표본공분산행렬
    - C_w: 전체 자료에 대한 표본공분산행렬
    - n_g: 그룹 g의 관측치 수
    - 검정통계량 $B = a\ function\ of \log(M) \sim \mathrm{X}^2(\frac{1}{2}p(p+1)(G-1))$
- R에서 공분산행렬의 등분산성 검정 해주는 패키지 없음
- BoxMTest function (by Ranjan Maitra, Iowa State University) 사용

```
> BoxMTest(mba[,1:2],mba[,3])
---------------------------------------------------
 MBox Chi-sqr. df P
---------------------------------------------------
   16.6653    16.0745           6       0.0134
---------------------------------------------------
Covariance matrices are significantly different.
$MBox
accepted
 16.6653

$ChiSq
accepted
16.07448

$df
[1] 6

$pValue
  accepted
0.01335976
```