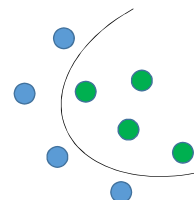
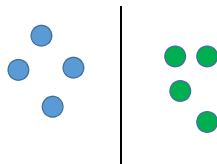


10.4 판별분석(Discriminant Analysis) II

- 이차판별분석 -

4. 선형판별분석 vs. 이차판별분석

LDA	QDA
분산-공분산 행렬이 범주 관계없이 동일한 경우	분산-공분산 행렬이 범주별로 다른 경우
(+) 적은 파라미터를 사용, 낮은 분산	(-) 많은 파라미터를 사용, 높은 분산
(-) 낮은 유연성	(+) 높은 유연성



5. 이차판별분석(QDA)

10.4 판별분석 II

• 모집단 등분산 검정

분산-공분산 행렬이 범주별로 다른 경우, 이차판별분석(QDA)을 실시



Box's M-test

귀무가설 : 모집단의 분산-공분산 행렬이 동일
대립가설 : 모집단의 분산-공분산 행렬이 동일 X

등분산검정을 위한 패키지 : biotools

```
install.packages("biotools")
library(biotools)
boxM(iris[1:4], iris$Species)

> boxM(iris[1:4], iris$Species)

Box's M-test for Homogeneity of Covariance Matrices

data: iris[1:4]
Chi-Sq (approx.) = 140.94, df = 20, p-value < 2.2e-16
```

p-value~0

귀무가설(등분산 가정)이 기각
→ QDA 실시!

5. 이차판별분석(QDA)

10.4 판별분석 II

• QDA 함수 : qda(종속변수~독립변수, data=학습 데이터 이름, prior=사전 확률)

```
# Quadratic Discriminant Analysis (QDA)
iris.qda <- qda(Species ~ ., data=train, prior=c(1/3,1/3,1/3))
iris.qda
```

```
> iris.qda
Call:
qda(Species ~ ., data = train, prior = c(1/3, 1/3, 1/3))

Prior probabilities of groups:
      setosa versicolor virginica
0.3333333  0.3333333  0.3333333

Group means:
      Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa           5.051613    3.461290    1.480645    0.2387097
versicolor       5.935484    2.745161    4.267742    1.3129032
virginica         6.634211    2.965789    5.597368    2.0289474
```

경우에 따라 다르게 줄수 있음

독립변수에 대한 그룹별 평균값

5. 이차판별분석(QDA)

10.4 판별분석 II

- 검증 데이터에 QDA 결과를 적용하여 범주 추정

```
# predict test data set n=50
testpredq <- predict(iris.qda, test)
```

```
> testpredq
$class
[1] setosa      setosa      setosa      setosa      setosa
[6] setosa      setosa      setosa      setosa      setosa
[11] setosa      setosa      setosa      setosa      setosa
[16] setosa      setosa      setosa      setosa      versicolor
[21] versicolor  versicolor  versicolor  versicolor  virginica
[26] versicolor  versicolor  versicolor  versicolor  versicolor
[31] versicolor  versicolor  versicolor  versicolor  versicolor
[36] versicolor  versicolor  versicolor  virginica   virginica
[41] virginica   virginica   virginica   virginica   virginica
[46] virginica   virginica   virginica   virginica   virginica
Levels: setosa versicolor virginica

$posterior
      setosa      versicolor      virginica
2      1.000000e+00  2.889417e-18  6.455011e-33
8      1.000000e+00  2.178894e-22  4.689120e-36
14     1.000000e+00  6.827538e-18  7.188376e-31
16     1.000000e+00  1.663584e-32  1.131649e-48
19     1.000000e+00  2.144558e-24  4.509102e-40
```

추정 범주

세 개 범주의 사후 확률(posterior probability)을 구한 후 max값의 범주로 할당

POSTECH
POSEIDON UNIVERSITY OF SCIENCE AND TECHNOLOGY

5

5. 이차판별분석(QDA)

10.4 판별분석 II

- 정확도 산정 : 오분류율 (검증데이터)

```
# accuracy of QDA
CrossTable(x=testLabels,y=testpredq$class, prop.chisq=FALSE)
```

Observations in Table: 50

testLabels	testpredq\$class			Row Total
	setosa	versicolor	virginica	
setosa	19	0	0	19
	1.000	0.000	0.000	0.380
	1.000	0.000	0.000	
	0.380	0.000	0.000	
versicolor	0	18	1	19
	0.000	0.947	0.053	0.380
	0.000	1.000	0.077	
	0.000	0.360	0.020	
virginica	0	0	12	12
	0.000	0.000	1.000	0.240
	0.000	0.000	0.923	
	0.000	0.000	0.240	
Column Total	19	18	13	50
	0.380	0.360	0.260	

- 정확도 : 49/50 -> 98%
- versicolor를 virginica로 잘못 예측
- 오분류율 : 1/50 -> 2%

POSTECH
POSEIDON UNIVERSITY OF SCIENCE AND TECHNOLOGY

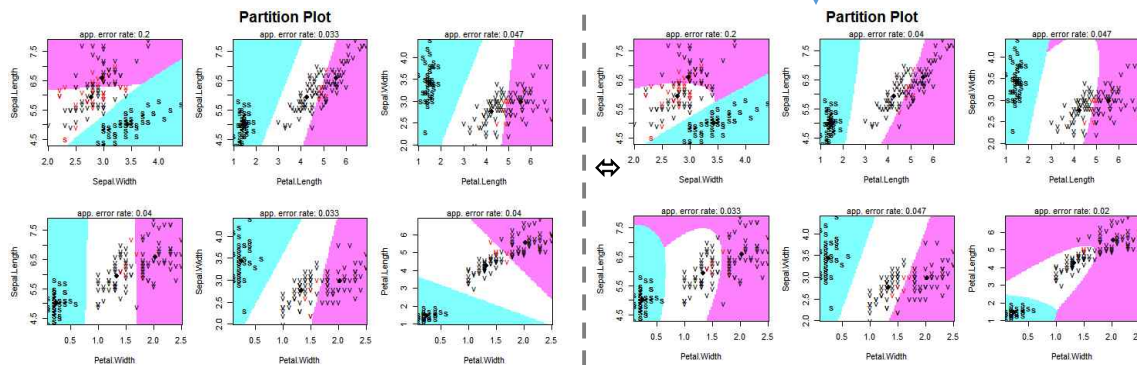
6

6. LDA vs. QDA (Partition Plot)

• Partition Plot : partimat()

```
install.packages("klaR")
library(klaR)
partimat(as.factor(iris$Species) ~ ., data=iris, method="lda")
```

```
partimat(as.factor(iris$Species) ~ ., data=iris, method="qda")
```



LDA

QDA

