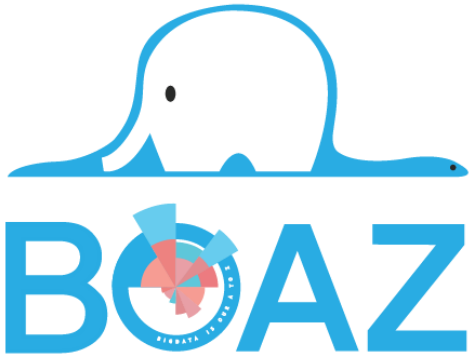
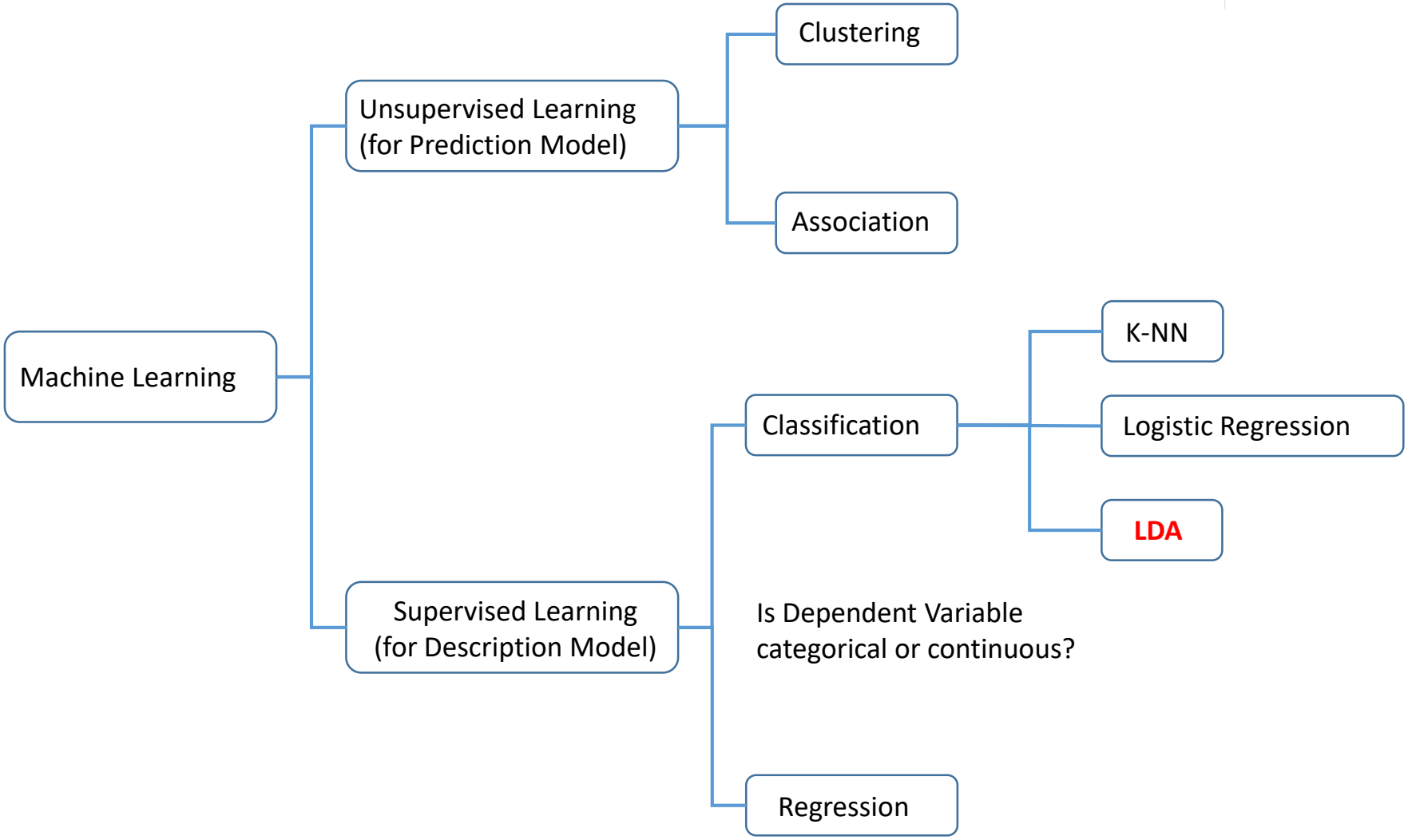


2016년 10월 06일

Linear Discriminant Analysis

F조: 나여영 이재혁 정회빈 최자연

About Machine Learning



Introduction to Regression

		Dependent Variables	
Independent Variables		Continuous	Categorical
	Continuous	Regression	Logistic Regression
	Continuous + Categorical	ANCOVA	
	Categorical	ANOVA	Chi-Square

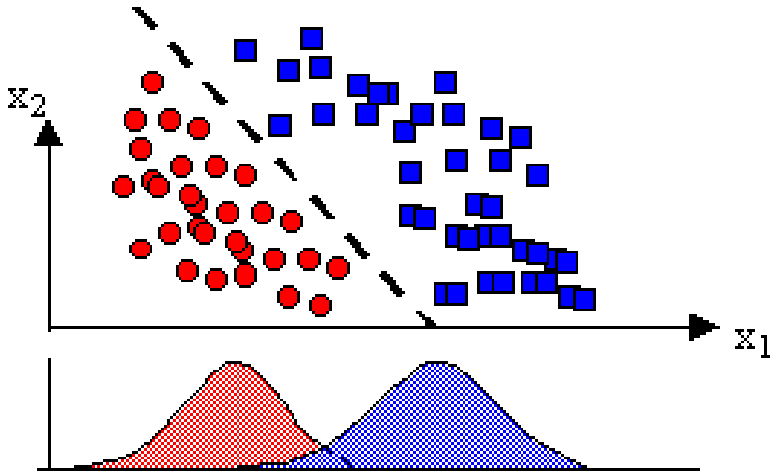
More About Classification

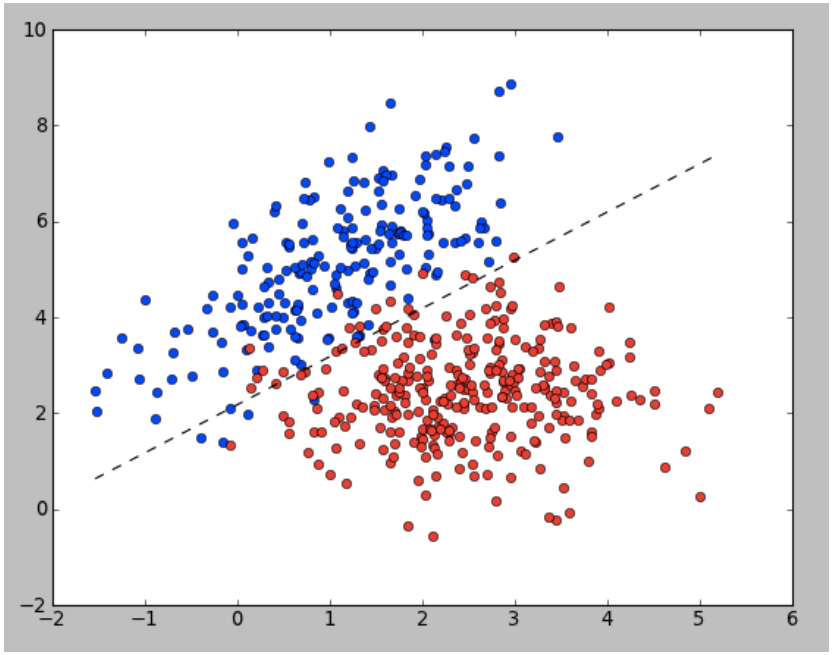
	Logistic Regression	LDA	QDA
Type of dependent variables	Categorical		
Assumption of Normal Distribution	X	O	O
Homogeneity of Covariance Matrix	X	O	X

LDA (Linear Discriminant Analysis)

목적 : 객체를 몇 개의 범주로 분류하기 위해 사용.

- 반응변수의 클래스 수가 2보다 클 때 일반적으로 사용한다.





LDA를 이용한 분류

1. 판별함수 (discriminant score)를 구한다
   - 판별함수 : $Y=dX$

2. 각 집단의 중심 위치 정하기
   - 각 집단 별 선형변화 $Y=d'X$의 평균값을 구한다.

3. 각 개체와 집단중심(center)과의 거리를 측정
   - 일반적으로 집단 중심을 평균으로 설정

4. 중심과의 거리가 가까운 집단으로 분류
   - 이 거리를 분류함수라고 한다.
   - 분류함수는 집단의 개수만큼 계산된다.

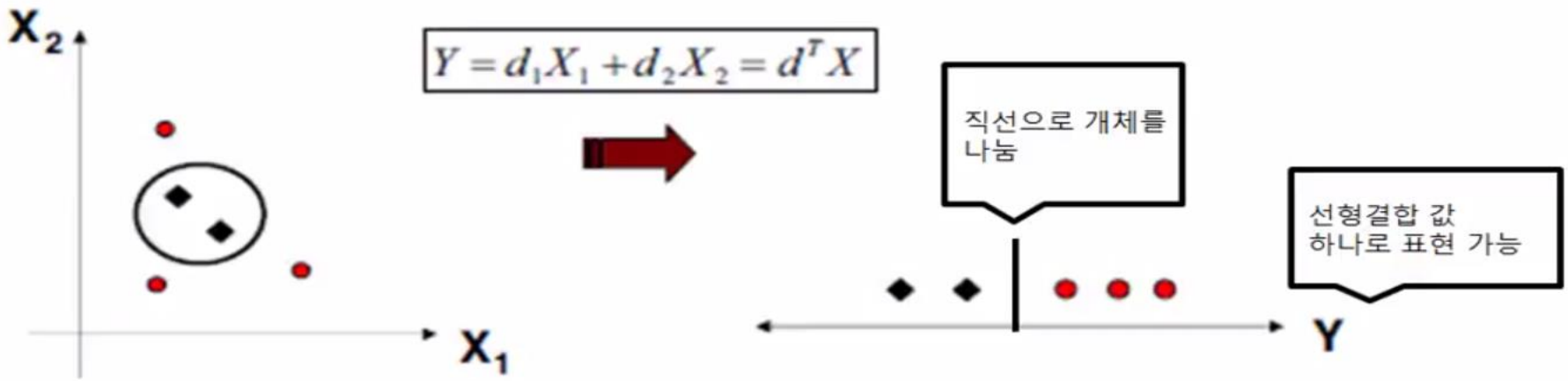
피셔 판별함수

$$E[x] = \begin{cases} \mu_1, & x\text{가 범주 1에 속할 때} \\ \mu_2, & x\text{가 범주 2에 속할 때} \end{cases}$$

$$Var[x] = \Sigma \qquad \text{(범주에 관계없이 동일)}$$

$$Z = \omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_p X_p = \omega^T x$$

$$\left(\lambda = \frac{\text{범주 간 } z\text{의 평균차이}}{z\text{의 분산}}\right)$$



피셔 판별함수

$$\lambda = \frac{\text{범주 간 } z\text{의 평균차이}}{z\text{의 분산}}$$

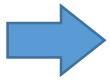
1) 범주간 Z의 평균차이 $= E[Z|\text{범주1}] - E[Z|\text{범주2}] = \omega^T\mu_1 - \omega^T\mu_2$

2) $Var[Z] = Var[\omega^T x] = \omega^T \sum \omega$

$$\lambda = \frac{\omega^T(\mu_1 - \mu_2)}{\omega^T \sum \omega}$$

피셔 판별함수

$$\lambda = \frac{\omega^T(\mu_1 - \mu_2)}{\omega^T \sum \omega}$$



ω로 미분후 = 0

$$\mu_1 - \mu_2 = \frac{2\omega^T(\mu_1 - \mu_2)}{\omega^T \sum \omega} \sum \omega$$

$$\omega \propto \sum^{-1} \omega^T(\mu_1 - \mu_2)$$

(판별계수 추정 : $\hat{\omega} = S_p{}^{-1}(\overline{x^{(1)}} - \overline{x^{(2)}})$ $\quad (S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2})$

▶ 피셔의 판별함수

$$Z = \omega^T x = (\mu_1 - \mu_2) \sum^{-1} x$$

피셔함수의 분류규칙

$|\hat{\omega}^T(x - \overline{x^{(1)}})| \leq |\hat{\omega}^T(x - \overline{x^{(2)}})|$ 이면, x를 범주1로 분류

$|\hat{\omega}^T(x - \overline{x^{(1)}})| > |\hat{\omega}^T(x - \overline{x^{(2)}})|$ 이면, x를 범주2로 분류

피셔함수의 분류규칙

$\overline{Z_1} = \hat{\omega}^T \overline{x^{(1)}} > \overline{Z_2} = \hat{\omega}^T \overline{x^{(2)}}$ 일 때,

$\hat{Z} = \hat{\omega}^T x \geq \bar{Z}$ 이면, x를 범주1로 분류

$\hat{Z} = \hat{\omega}^T x < \bar{Z}$ 이면, x를 범주2로 분류

$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{n_1\overline{Z_1} + n_2\overline{Z_2}}{n_1 + n_2}$

▶ 피셔의 판별함수에 따른 분류 경계식

$$\widehat{\omega^T} x = \bar{Z}$$

피셔함수의 분류규칙

▶ 마할라노비스의 거리 정의

$$d^2(x_1, x_2) = (x_1 - x_2)^T \sum^{-1} (x_1 - x_2)$$

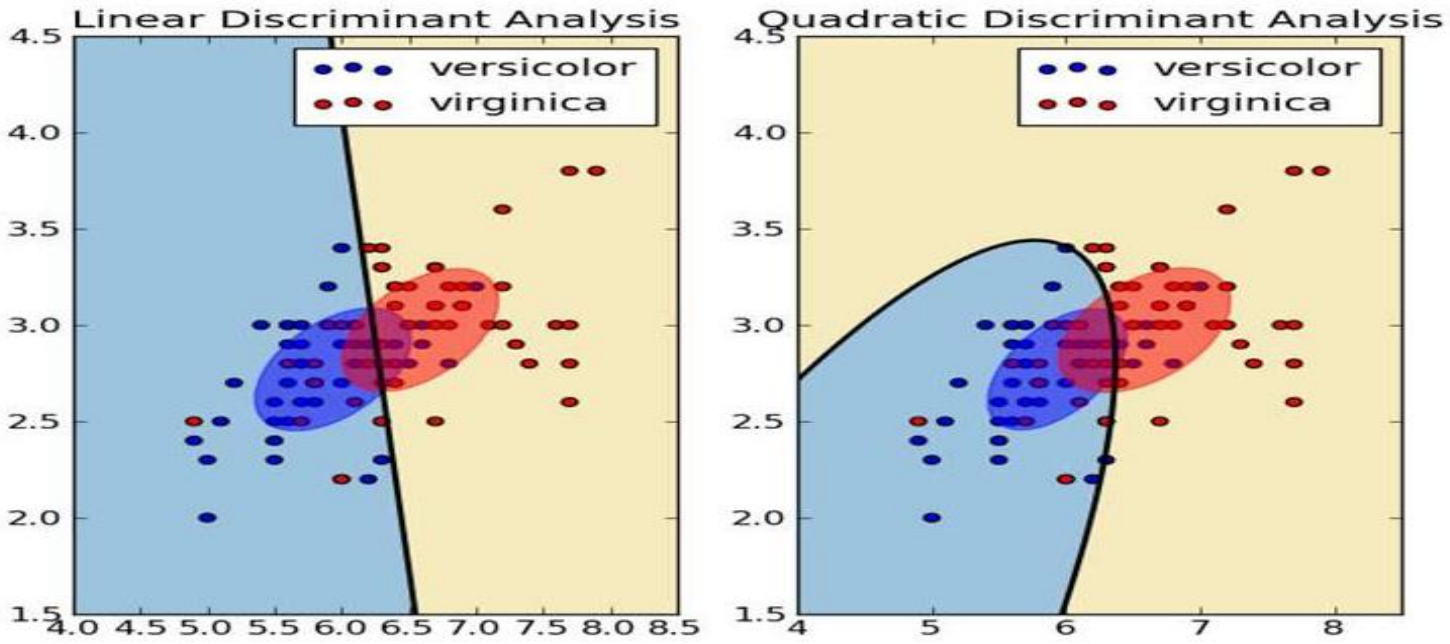
▶ 분류 규칙

$d^2(x, \mu_1) \leq d^2(x, \mu_2)$이면 , x를 범주 1로 분류

LDA vs QDA

- LDA
  : 정규분포의 분산 – 공분산 행렬이 범주에 상관없이 동일하다 가정

- QDA
  : 정규분포의 분산 – 공분산 행렬이 범주별로 다르다 가정



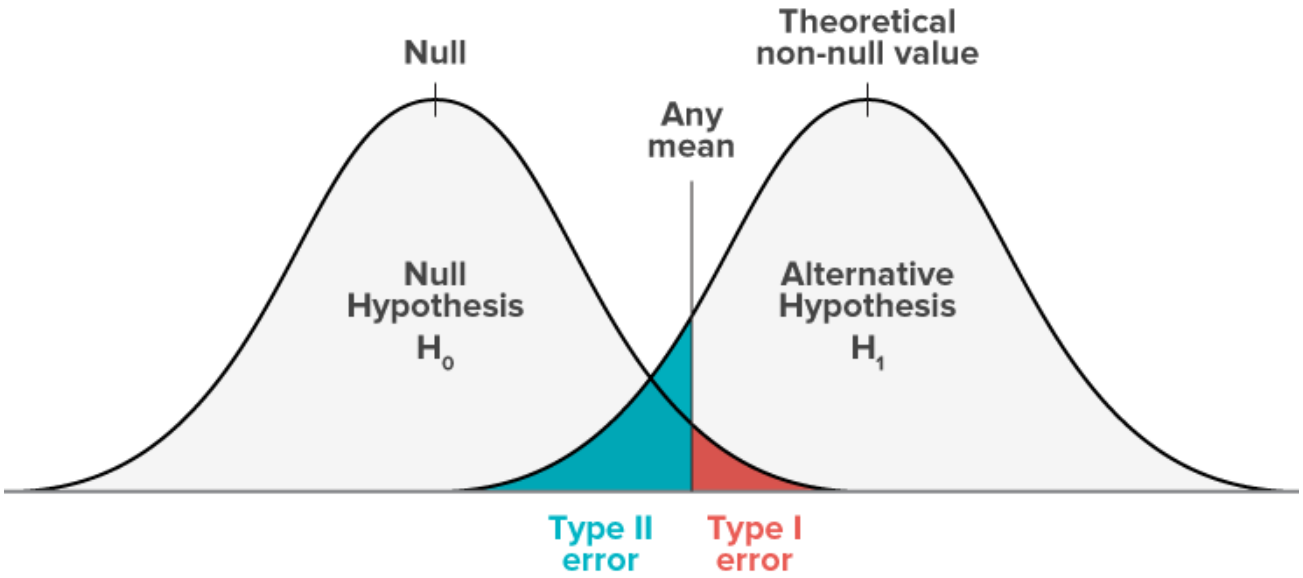
1. Error(오류)–통계학 입문

		reality	
		H_0 = true	H_0 = false
conclusion	H_0 = true	OK	type II error
	H_0 = false	type I error	OK

TEXT

type I error(α)와 type II error(β)는 서로 trade-off 관계

->보통 α를 고정시키고 β 최소화 시키는 방법 사용



과적합(overfitting)??

Error

혼동 행렬 (confusing matrix)		실제 연체 상태		
		아니오	예	합계
예측한 연체 상태	아니오	9644	252	9896
	예	23	81	104
	합계	9667	333	10000

error rate = ??

23/9667=0.238%

252/333=75.7%

sensitivity, specificity

혼동 행렬 (confusing matrix)		실제 연체 상태		
		아니오	예	합계
예측한 연체 상태	아니오	9644	252	9896
	예	23	81	104
	합계	9667	333	10000

sensitivity = P(예측연체상태=예|실제연체상태=예)
specificity = P(예측연체상태=아니오|실제연체상태=아니오)

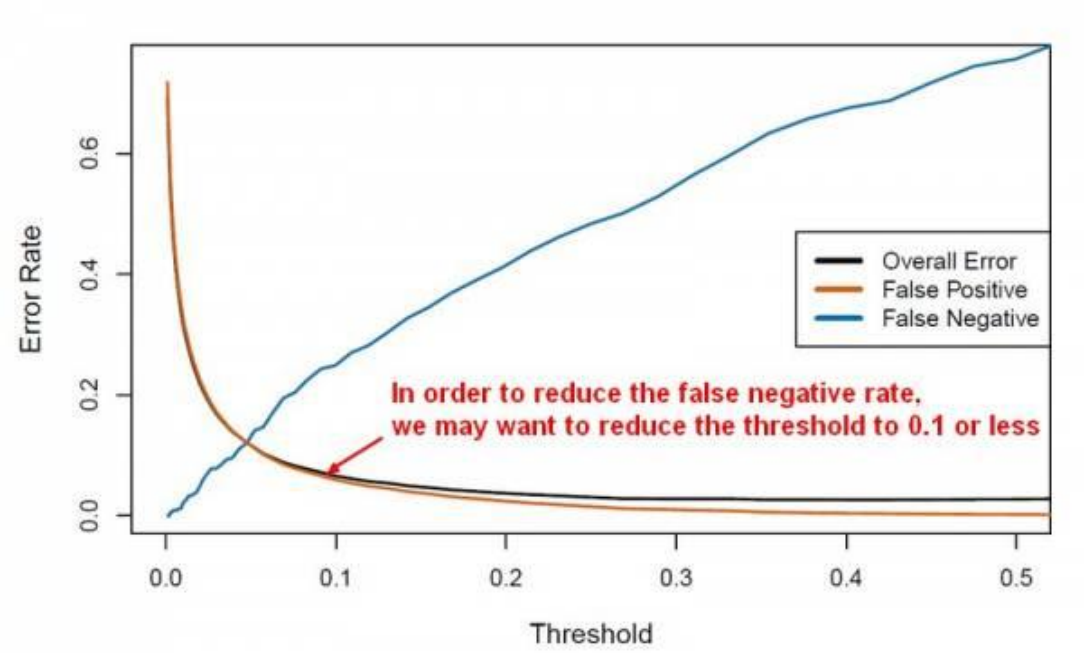
sensitivity=81/333=24.3%
specificity=9644/9667=99.8%

sensitivity, specificity

혼동 행렬 (confusing matrix)		실제 연체 상태		
		아니오	예	합계
예측한 연체 상태	아니오	True Neg.	False Pos.	N
	예	False Neg.	True Pos.	P
	합계	N*	P*	

FN/N => Type I error, 1-specificity
TP/P => 1-Type II error, power, sensitivity

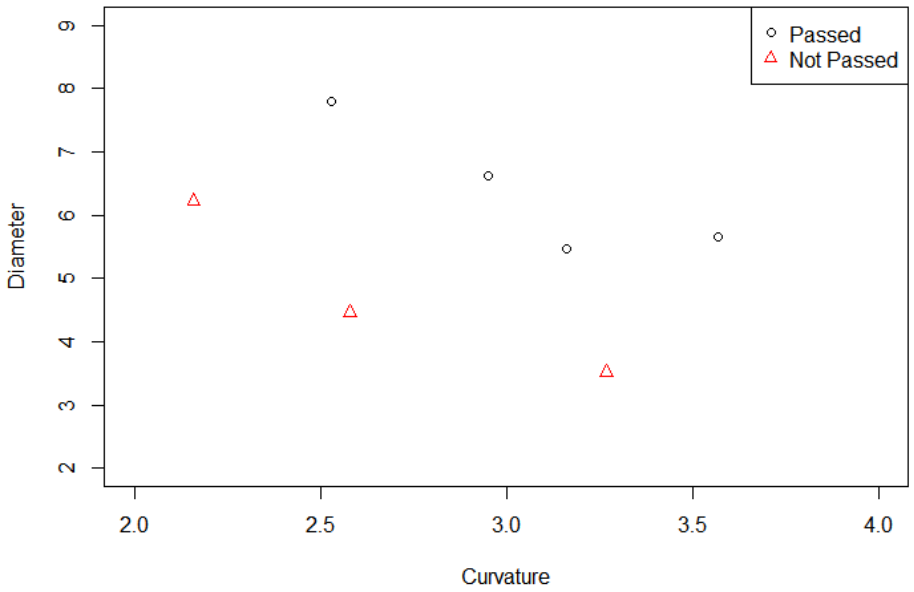
ROC curve



AUC=ROC curve 아래의 면적으로, 분류기 성능 지표

Numerical Example

Curvature	Diameter	Quality Control Result
2.95	6.63	Passed
2.53	7.79	Passed
3.57	5.65	Passed
3.16	5.47	Passed
2.58	4.46	Not Passed
2.16	6.22	Not Passed
3.27	3.52	Not Passed



http://people.revoledu.com/kardi/tutorial/LDA/LDA%20Formula.htm

Numerical Example

$$X = \begin{bmatrix} 2.95 & 6.63 \\ 2.53 & 7.79 \\ 3.57 & 5.65 \\ 3.16 & 5.47 \\ 2.58 & 4.46 \\ 2.16 & 6.22 \\ 3.27 & 3.52 \end{bmatrix}$$

$$\mu = [2.889 \quad 5.677]$$

$$X_1 = \begin{bmatrix} 2.95 & 6.63 \\ 2.53 & 7.79 \\ 3.57 & 5.65 \\ 3.16 & 5.47 \end{bmatrix}$$

$$\mu_1 = [3.053 \quad 6.385]$$

$$X_2 = \begin{bmatrix} 2.58 & 4.46 \\ 2.16 & 6.22 \\ 3.27 & 3.52 \end{bmatrix}$$

$$\mu_2 = [2.67 \quad 4.733]$$

$$y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

Numerical Example

$$C_1 = \frac{1}{n_1 - 1}(X_1 - 1_{4\times1}\mu)^T(X_1 - 1_{4\times1}\mu)$$

$$= \frac{1}{3}\left[\begin{bmatrix}2.95 & 6.63\\ 2.53 & 7.79\\ 3.57 & 5.65\\ 3.16 & 5.47\end{bmatrix} - \begin{bmatrix}2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\end{bmatrix}\right]^T\left[\begin{bmatrix}2.95 & 6.63\\ 2.53 & 7.79\\ 3.57 & 5.65\\ 3.16 & 5.47\end{bmatrix} - \begin{bmatrix}2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\end{bmatrix}\right] = \begin{bmatrix}0.223 & -0.258\\ -0.258 & 1.806\end{bmatrix}$$

$$C_2 = \frac{1}{n_2 - 1}(X_2 - 1_{3\times1}\mu)^T(X_2 - 1_{3\times1}\mu)$$

$$= \frac{1}{2}\left[\begin{bmatrix}2.58 & 4.46\\ 2.16 & 6.22\\ 3.27 & 3.52\end{bmatrix} - \begin{bmatrix}2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\end{bmatrix}\right]^T\left[\begin{bmatrix}2.58 & 4.46\\ 2.16 & 6.22\\ 3.27 & 3.52\end{bmatrix} - \begin{bmatrix}2.889 & 5.677\\ 2.889 & 5.677\\ 2.889 & 5.677\end{bmatrix}\right] = \begin{bmatrix}0.223 & -0.258\\ -0.258 & 1.806\end{bmatrix}$$

$$C = \frac{(n_1 - 1)C_1 + (n_2 - 1)C_2}{n_1 + n_2 - 2} = \begin{bmatrix}0.288 & -0.323\\ -0.323 & 2.369\end{bmatrix}$$

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Numerical Example – 피셔의 판별함수

$$W = C^{-1}(\mu_1 - \mu_2)^T = \begin{bmatrix} 4.099 & 0.599 \\ 0.599 & 0.498 \end{bmatrix} \begin{bmatrix} 0.383 \\ 1.652 \end{bmatrix} = \begin{bmatrix} 2.493 \\ 1.037 \end{bmatrix}$$

판별계수 추정

$$\hat{\omega} = {S_p}^{-1}(\overline{x^{(1)}} - \overline{x^{(2)}})$$

$$Z_1 = W^T {\mu_1}^T = [2.493 \quad 1.037] \begin{bmatrix} 3.053 \\ 6.385 \end{bmatrix} = 14.232$$

$$Z_2 = W^T {\mu_2}^T = [2.493 \quad 1.037] \begin{bmatrix} 2.67 \\ 4.733 \end{bmatrix} = 11.564$$

$$Z = \frac{n_1}{n_1 + n_2} Z_1 + \frac{n_2}{n_1 + n_2} Z_2 = 13.089$$

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{n_1 \overline{Z_1} + n_2 \overline{Z_2}}{n_1 + n_2}$$

$$Z = W^T X = [2.493 \quad 1.037] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = 2.493X_1 + 1.037X_2 = 13.089$$

$$X_2 = -\frac{2.493}{1.037} X_1 + \frac{13.089}{1.037} = -2.404X_1 + 12.622$$

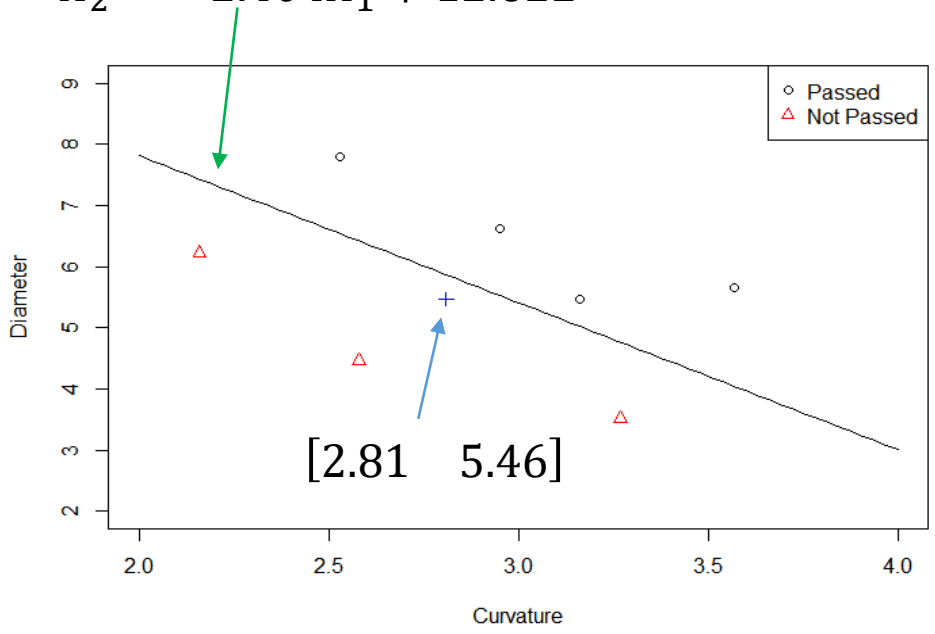
결정경계(Decision Boundary)

Numerical Example – 피셔의 판별함수

$$|W^T(x-\mu_1)| = [2.493 \quad 1.037]\begin{bmatrix}-0.243\\-0.925\end{bmatrix} = 1.565$$

$$|W^T(x-\mu_2)| = [2.493 \quad 1.037]\begin{bmatrix}0.140\\0.727\end{bmatrix} = 1.103$$

$$X_2 = -2.404X_1 + 12.622$$



$|\hat{\omega}^T(x-\overline{x^{(1)}})| \leq |\hat{\omega}^T(x-\overline{x^{(2)}})|$

x를 범주1로 분류

$|\hat{\omega}^T(x-\overline{x^{(1)}})| > |\hat{\omega}^T(x-\overline{x^{(2)}})|$

x를 범주2로 분류

$x = [2.81 \quad 5.46]$

x는 범주 2에 속함

Numerical Example – ISLR 교재에 나온 방법

$$f_i = \mu_i C^{-1} x^T - \frac{1}{2}\mu_i C^{-1} {\mu_i}^T + \ln(\frac{n_i}{n_i + n_j})$$

$f_i > f_j$ 이면 x가 그룹 i에 속하는 것으로 분류

$$f_1 = \mu_1 C^{-1} x^T - \frac{1}{2}\mu_1 C^{-1} {\mu_1}^T + \ln\left(\frac{n_1}{n_1 + n_2}\right) = 31.166$$

$$f_2 = \mu_2 C^{-1} x^T - \frac{1}{2}\mu_2 C^{-1} {\mu_2}^T + \ln\left(\frac{n_2}{n_1 + n_2}\right) = 31.11$$

$f_1 > f_2$ 이므로 x는 그룹 1에 속함

표본크기가 너무 작아서 결과가 다르게 나온 것으로 보임