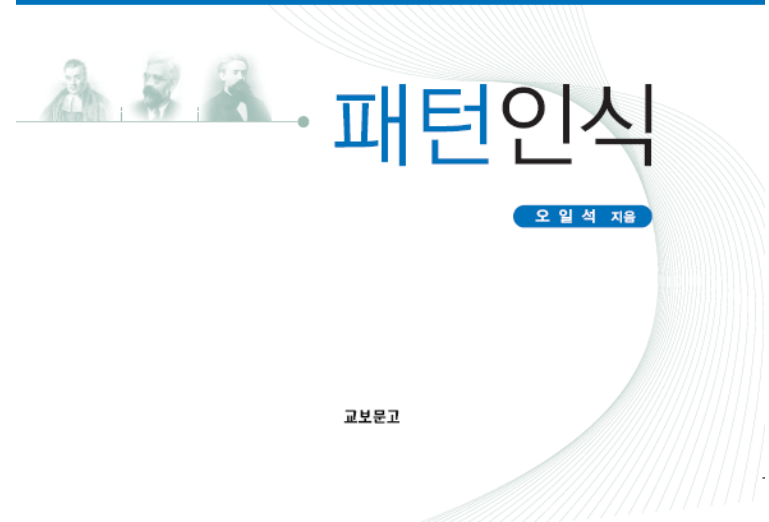
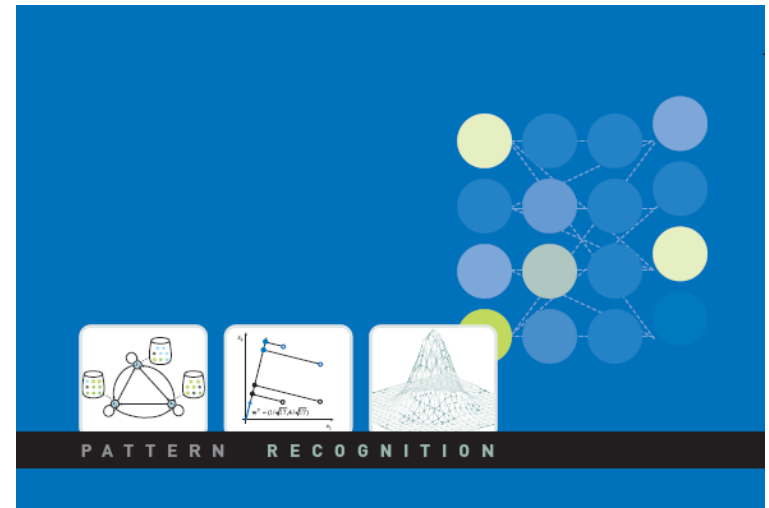


8장. 특징 추출

오일석, 패턴인식, 교보문고, 2008.



들어가는 말

- 특징 추출의 예
 - 필기 숫자 인식

6

- 크기 정규화
- 이진화

00001100
00010000
00100000
01100000
11000110
11000011
11000001
11111110

때때로
이진
화

방법 1: 화소 각각을 특징으로 삼음
64 차원 특징 벡터
 $\mathbf{x}=(0,0,0,0,1,1,0,0,0,0,0,1,\dots,1,1,0)^T$

방법 2: 가로
이등분과 세로
이등분하여 검
은 화소의 개
수 비율을 특
징으로 삼음

$x_2=14/10$
 $\frac{14}{10}$
00001100
00010000
00100000
01100000
11000110
11000011
11000001
11111110

6
18
 $x_1=6/18$

$\mathbf{x}=(6/18,14/10)^T$

- 특징의 우수성 기준
 - 분별력
 - 차원

또 다른 방법으로는?

8.1.1 실제 세계의 다양성

- 특징 생성
 - 외부의 물리적 패턴을 특징 벡터라는 수학적 표현으로 변환

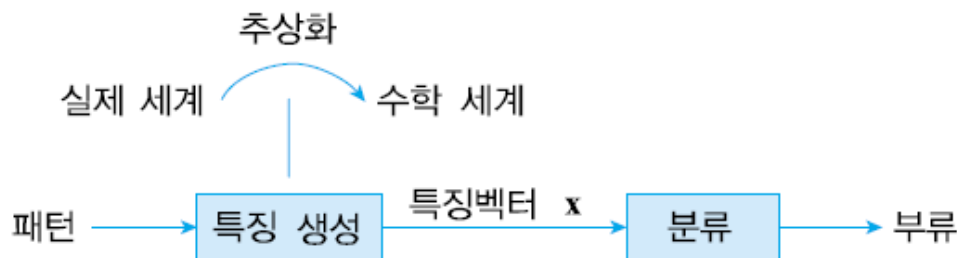


그림 8.1 패턴 인식의 일반적인 틀

- 특징 생성 과정은 매우 다양
 - 특징 추출은 외부 환경에 맞게 설계해야 하기 때문
 - 숫자와 한글은 다른 특징 필요할 수 있음
 - 한글도 통째로 인식하는 방법과 자소로 분할한 후 인식하는 방법이 다른 특징 필요할 수 있음
 - 정면 얼굴로 국한하는 하는 경우와 제약이 없는 얼굴 인식의 특징은 다를 수 있음

8.1.2 특징 추출과 특징 선택

- 센싱으로 얻은 신호의 다양성
 - 영상
 - 시간성 신호
 - 측정 벡터
- 특징 추출과 특징 선택

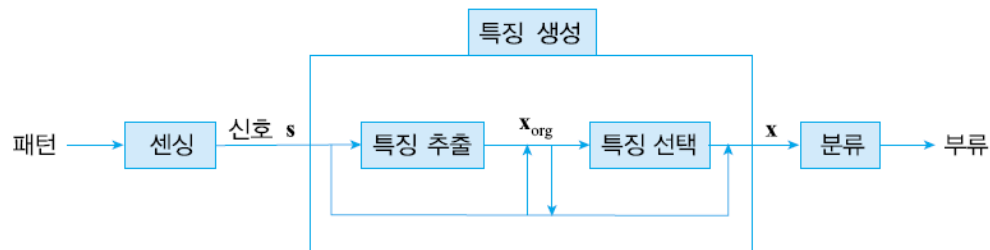


그림 8.2 특징 생성의 절차

$$\left. \begin{array}{l} \text{특징추출} \quad \mathbf{X}_{org} = e(\mathbf{s}) \\ \text{특징선택} \quad \mathbf{X} = s(\mathbf{X}_{org}) \end{array} \right\}$$

- 다양한 상황

$$\left. \begin{array}{l} \mathbf{X} = \mathbf{s} \\ \mathbf{X} = e(\mathbf{s}) \\ \mathbf{X} = s(\mathbf{s}) \\ \mathbf{X} = s(e(\mathbf{s})) \end{array} \right\}$$

8.2 영역에서 특징 추출

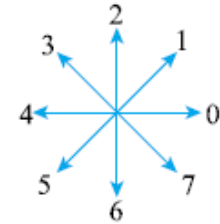
■ 영역의 표현

	0	1	2	3	4	5
0						1
1					1	1
2	1				1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	
5		1	1	1		

(a) 이진 배열 표현

$R = \{(5,0), (4,1), (5,1), (0,2), (4,2), (5,2),$
 $(0,3), (1,3), (2,3), (3,3), (4,3), (5,3),$
 $(0,4), (1,4), (2,4), (3,4), (4,4), (1,5),$
 $(2,5), (3,5)\}$

(b) 집합 표현



$C = (0,4) \cdot 2 \cdot 2 \cdot 7 \cdot 0 \cdot 0 \cdot 1 \cdot 2 \cdot 1 \cdot 6 \cdot 6 \cdot 6 \cdot 5 \cdot 5 \cdot 4 \cdot 4 \cdot 3$

(c) 체인 코드 표현

그림 8.3 영역의 표현

8.2.1 모양에 관련한 특징

- 모멘트와 중심 모멘트

$$m_{pq} = \sum_{(x,y) \in R} x^p y^q$$

$$\bar{m}_{pq} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q$$

- 여러 가지 특징들

면적 $a = m_{00}$

중점 $(\bar{x}, \bar{y}) = \left(\frac{m_{10}}{a}, \frac{m_{01}}{a} \right)$

행 분산 $v_{rr} = \frac{\bar{m}_{20}}{a}$

열 분산 $v_{cc} = \frac{\bar{m}_{02}}{a}$

혼합 분산 $v_{rc} = \frac{\bar{m}_{11}}{a}$

둘레 $p = n_{\text{even}} + n_{\text{odd}} \sqrt{2}$
이때 n_{even} = 짝수를 갖는 체인의 개수
 n_{odd} = 홀수를 갖는 체인의 개수

둥근 정도 $r = \frac{4\pi a}{p^2}$

- 불변 특성

- 이동 불변, 크기 불변, 회전 불변

8.2.1 모양에 관련한 특징

■ 예제 8.1 모양 특징의 추출

1번 영역의 특징

면적 $a = 20$

$$\text{중점 } (\bar{x}, \bar{y}) = \left(\frac{\sum_{(x,y) \in R} x}{20}, \frac{\sum_{(x,y) \in R} y}{20} \right) = (x\text{의 합}/20, y\text{의 합}/20) = (2.7, 3.05)$$

$$\text{행 분산 } v_{rr} = \frac{\sum_{(x,y) \in R} (x - 2.7)^2}{20} = 3.01$$

$$\text{열 분산 } v_{cc} = \frac{\sum_{(x,y) \in R} (y - 3.05)^2}{20} = 1.848$$

$$\text{혼합 분산 } v_{rc} = \frac{\sum_{(x,y) \in R} (x - 2.7)(y - 3.05)}{20} = -1.135$$

$$\text{둘레 } p = 10 + 6\sqrt{2} = 18.485$$

$$\text{둥근 정도 } r = \frac{4\pi * 20}{18.485^2} = 0.736$$

	0	1	2	3	4	5	6	7	8	9	0	1
0						1						
1					1	1						
2	1				1	1				3	3	
3	1	1	1	1	1	1				3	3	
4	1	1	1	1	1					3	3	
5		1	1	1						3	3	
6										3	3	
7					2	2				3	3	
8				2	2	2	2			3	3	
9				2	2	2	2			3	3	
0					2	2				3	3	
1												

그림 8.5 세 개의 영역

	면적 a	중점 (\bar{x}, \bar{y})	행 분산 v_{rr}	열 분산 v_{cc}	혼합 분산 v_{rc}	둘레 p	둥근 정도 r
영역 1	20	(2.7, 3.05)	3.01	1.848	-1.135	18.485	0.736
영역 2	12	(4.5, 8.5)	0.917	0.917	0.0	9.657	1.617
영역 3	18	(9.5, 6)	0.25	6.667	0.0	18	0.698

8.2.2 투영과 프로파일 특징

■ 투영 특징

행 투영 h_i = 행 i 에 있는 1의 개수, $0 \leq i \leq N-1$ (8.12)

열 투영 v_j = 열 j 에 있는 1의 개수, $0 \leq j \leq M-1$ (8.13)

- $N+M$ 차원의 특징 벡터를 얻게 됨
- 분별력은?

■ 예제 8.2

	0	1	2	3	4	5	
0						1	1
1					1	1	2
2	1				1	1	3 앞
3	1	1	1	1	1	1	6 중간 0요
4	1	1	1	1	1		5
5		1	1	1			3
	3	3	3	3	4	4	

열 투영

그림 8.6 투영 특징

특징 벡터 $\mathbf{x} = (3, 3, 3, 3, 4, 4, 1, 2, 3, 6, 5, 3)^T$

이웃한 두 개의 투영을 평균한다면 6 차원 특징 벡터 $\mathbf{x} = (3, 3, 4, 1.5, 4.5, 4)^T$

8.2.2 투영과 프로파일 특징

■ 프로파일 특징

- $2(N+M)$ 차원의 특징 벡터를 얻게 됨

상 프로파일 t_j = 위에서 열 j 를 바라보았을 때 처음 1까지의 거리,

$$0 \leq j \leq M-1 \quad (8.14)$$

우 프로파일 r_i = 오른쪽에서 행 i 를 바라보았을 때 처음 1까지의 거리,

$$0 \leq i \leq N-1 \quad (8.15)$$

하 프로파일 b_j = 아래에서 열 j 를 바라보았을 때 처음 1까지의 거리,

$$0 \leq j \leq M-1 \quad (8.16)$$

좌 프로파일 l_i = 왼쪽에서 행 i 를 바라보았을 때 처음 1까지의 거리,

$$0 \leq i \leq N-1 \quad (8.17)$$

■ 예제 8.3

		상 프로파일								
		2	3	3	3	1	0			
상 프로파일	5						1	0	하 프로파일	
	4					1	1	0		
	0	1				1	1	0		
	0	1	1	1	1	1	1	0		
	0	1	1	1	1	1		1		
	1		1	1	1			2		
		1	0	0	0	1	2			
		하 프로파일								

그림 8.7 프로파일 특징

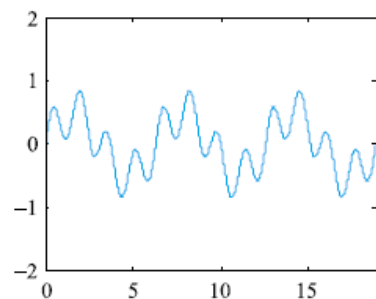
특징 벡터 $\mathbf{x} = (2, 3, 3, 3, 1, 0, 0, 0, 0, 0, 1, 2, 2, 1, 0, 0, 0, 1, 1, 0, 0, 0, 4, 5)^T$

이웃한 세 개의 값을 평균한다면 8 차원 특징 벡터 $\mathbf{x} = (2.66, 1.33, 0, 1, 1, 0.33, 0.33, 3)^T$

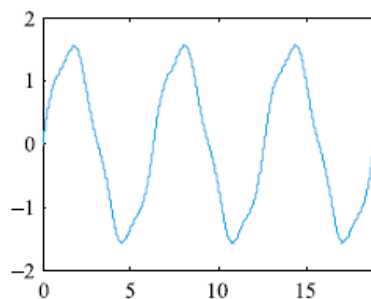
8.3 변환을 이용한 특징

■ 파형 신호에서 특징 추출

- 예) 지진파, 기계 진동파, 수중파, 음파, 재정 자립도 추이 곡선, 주식 곡선 등
- 파형에서 어떻게 특징을 추출할 것인가?
 - 파형은 기저 함수의 선형 결합으로 표현 가능
 - 선형 결합의 계수를 특징으로 취함. 계수를 어떻게 구할 것인가?



(a) 신호 $s_1(x) = 0.5g_1(x) + 1.5g_2(x)$



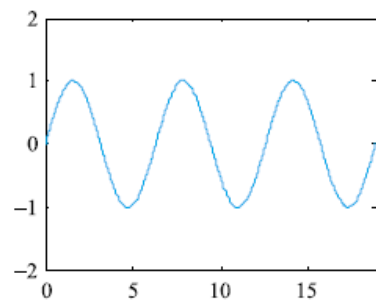
(b) 신호 $s_2(x) = 1.5g_1(x) + 0.5g_2(x)$

$$s_1(x) = 0.5g_1(x) + 1.5g_2(x)$$

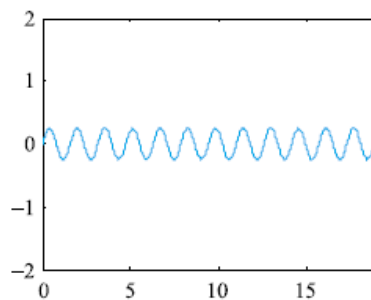
$$s_2(x) = 1.5g_1(x) + 0.5g_2(x)$$

신호 s_1 의 특징 벡터 $= (0.5, 1.5)^T$

신호 s_2 의 특징 벡터 $= (1.5, 0.5)^T$



(c) 기저 함수 $g_1(x) = \sin(x)$



(d) 기저 함수 $g_2(x) = 0.25\sin(4x)$

그림 8.8 입력 신호는 기저 함수의 선형 결합으로 표현할 수 있음

8.3.1 푸리에 변환

■ 이산 푸리에 변환

- 신호를 $\mathbf{s}=(s(0),s(1),\dots,s(n-1))^T$ 로 표현할 때,
- 푸리에 변환 $f(u)$ 는 시간 공간을 주파수 공간으로 바꾸어 준다.

$$f(u) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} s(i) \exp(-j \frac{2\pi ui}{n}), u = 0, \dots, n-1 \quad (8.19)$$

■ 예제 8.4 이산 푸리에 변환

- $\mathbf{s}=(1.7,0.6,1.5,1.2)^T$

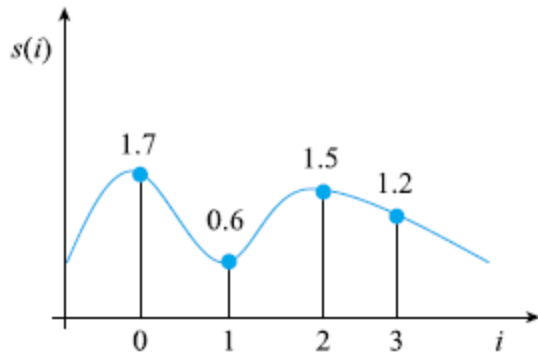


그림 8.9 신호 샘플링

$$f(0) = 0.5(1.7e^0 + 0.6e^0 + 1.5e^0 + 1.2e^0) = 2.5 + j0$$

$$f(1) = 0.5(1.7e^0 + 0.6e^{-j\pi/2} + 1.5e^{-j2\pi/2} + 1.2e^{-j3\pi/2})$$

$$= 0.5(1.7 + 0.6(\cos(-\pi/2) + j\sin(-\pi/2)) + 1.5(\cos(-2\pi/2)$$

$$+ j\sin(-2\pi/2)) + 1.2(\cos(-3\pi/2) + j\sin(-3\pi/2)))$$

$$= 0.5((1.7 + 0.6\cos(-\pi/2) + 1.5\cos(-2\pi/2) + 1.2\cos(-3\pi/2))$$

$$+ j(0.6\sin(-\pi/2) + 1.5\sin(-2\pi/2) + 1.2\sin(-3\pi/2))) = 0.1 + j0.3$$

$$f(2) = 0.5(1.7e^0 + 0.6e^{-j2\pi/2} + 1.5e^{-j4\pi/2} + 1.2e^{-j6\pi/2}) = 0.7 + j0$$

$$f(3) = 0.5(1.7e^0 + 0.6e^{-j3\pi/2} + 1.5e^{-j6\pi/2} + 1.2e^{-j9\pi/2}) = 0.1 - j0.3$$

u	실수부	허수부
0	2.5	0
1	0.1	0.3
2	0.7	0
3	0.1	-0.3

8.3.1 푸리에 변환

- 푸리에 특징

- 파워 스펙트럼의 값을 특징으로 취함

$$\text{파워 스펙트럼 } p(u) = \sqrt{\text{real}(f(u))^2 + \text{imag}(f(u))^2}, u = 0, \dots, n-1 \quad (8.20)$$

- 예제 8.5 푸리에 특징

$$p(0) = (2.5^2 + 0^2)^{1/2} = 2.5$$

$$p(1) = (0.1^2 + 0.3^2)^{1/2} = 0.316$$

$$p(2) = (0.7^2 + 0^2)^{1/2} = 0.7$$

$$p(3) = (0.1^2 + (-0.3)^2)^{1/2} = 0.316$$

즉 아래와 같은 4 차원 특징 벡터를 얻게 된다.

$$\mathbf{x} = (2.5, 0.316, 0.7, 0.316)^T$$

8.3.1 푸리에 변환

■ 푸리에 특징 추출 알고리즘

- d 개의 특징을 추출하고자 하면 $p(0), p(1), \dots, p(d-1)$ 을 취한다.
- 왜냐하면 u 가 커질수록 $p(u)$ 는 급속도로 작아지므로

알고리즘 [8.1]

푸리에 특징 추출

입력: 신호 s , 필요한 특징 개수 d

출력: d 차원의 특징 벡터 \mathbf{x}

알고리즘:

1. FFT 알고리즘을 이용하여 s 의 푸리에 변환을 구한다.
2. (8.20)으로 파워 스펙트럼 $p(u)$, $u=0, \dots, n-1$ 을 구한다.
3. $\mathbf{x} = (p(0), p(1), \dots, p(d-1))^T$;

■ 2 차원 푸리에 변환

$$f(u, v) = \frac{1}{n} \sum_{i=0}^{n-1} \sum_{k=0}^{n-1} s(i, k) \exp(-j \frac{2\pi(ui + vk)}{n}), u = 0, \dots, n-1, v = 0, \dots, n-1 \quad (8.21)$$

$$\text{파워 스펙트럼 } p(u, v) = \sqrt{\text{real}(f(u, v))^2 + \text{imag}(f(u, v))^2} \quad (8.22)$$

8.3.2 푸리에 기술자

■ 푸리에 기술자

- $n*n$ 영상에 푸리에 변환 적용하여 특징 추출
- 영역 경계 상의 점을 복소수로 표현한 뒤에 2 차원 푸리에 변환 적용

그림 8.3(a)에 있는 영역의 경계 표현:

$(0,4)-(0,3)-(0,2)-(1,3)-(2,3)-(3,3)-(4,2)-(4,1)-(5,0)-(5,1)-(5,2)-$
 $(5,3)-(4,4)-(3,5)-(2,5)-(1,5)$

$$\alpha(i) = x_i + jy_i, \quad 0 \leq i \leq n-1$$

푸리에 기술자
$$f(u) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \alpha(i) \exp(-j \frac{2\pi ui}{n}), \quad u = 0, \dots, n-1 \quad (8.23)$$

Joseph Fourier

(1768년 3월 21일 ~ 1830년 5월 16일) 프랑스

“모든 함수는 서로 다른 주파수를 갖는 수 많은 sine 파와 cosine 파의 합으로 표현할 수 있다.” 이것이 Fourier가 세상에 내놓은 아이디어이다. 이 아이디어를 기반으로 탄생한 푸리에 변환은 신호 처리와 영상 처리에 폭 넓게 쓰인다. 패턴 인식에서는 특징 추출을 위해 사용한다. Fourier의 이 아이디어는 그 당시 유명한 수학자들의 논쟁을 불러일으켰는데 Lagrange와 Laplace는 부정적으로 평가했다고 전한다.



Fourier는 대기의 가스가 지구의 온도를 높일 수 있다는 사실을 발견하기도 하였다. 그의 이 생각은 현재 지구가 앓고 있는 지구 온난화와 관련이 있다. Fourier의 정치적인 이력이 매우 흥미롭다. 그는 프랑스 혁명에 가담하였으며 파벌 싸움의 와중에 단두대에 오를 뻔한 위기도 맞았다. 이후 나폴레옹의 이집트 원정에 과학 자문관으로 동행하였고 나중에 나폴레옹의 요청으로 Grenoble 지역의 주지사를 역임하기도 하였다. 이후 나폴레옹과 심한 갈등을 겪기도 한다. Fourier에 대한 보다 상세한 내용은 [Arago05]를 참고하라.

[Arago05] Francois Arago, *Biographies of Distinguished Scientific Men*, (Translated by W. H. Smyth, Baden Powell, and Robert Grant) (Released on 2005 by the Project Gutenberg Ebook, <http://www.gutenberg.org/etext/16775>).

8.5 주성분 분석

- 주성분 분석 principal component analysis
 - 훈련 집합을 이용하여 매개 변수를 추정하고 그것을 이용하여 특징 추출함
 - 정보 손실을 최소화하는 조건에서 차원 축소
 - Karhunen-Loeve (KL) 변환 또는 Hotelling 변환이라고도 부름

8.5.1 동기

- 주성분 분석의 동기
 - \mathbf{U} 는 ‘정보 손실을 최소화하며’ 신호 \mathbf{s} 를 보다 낮은 차원의 특징 벡터 \mathbf{x} 로 변환 ($d < D$)
 - 변환 행렬 \mathbf{U} 는 $d \times D$ 행렬 (신호 \mathbf{s} 는 D 차원, 특징 벡터 \mathbf{x} 는 d 차원)
 - 두가지 문제
 - 차원 축소를 어떻게 표현할 것인가?
 - 정보 손실을 어떻게 수량화할 것인가?

$$\mathbf{x} = e(\mathbf{s}; \mathbf{U}) = \mathbf{U}\mathbf{s} \quad (8.26)$$

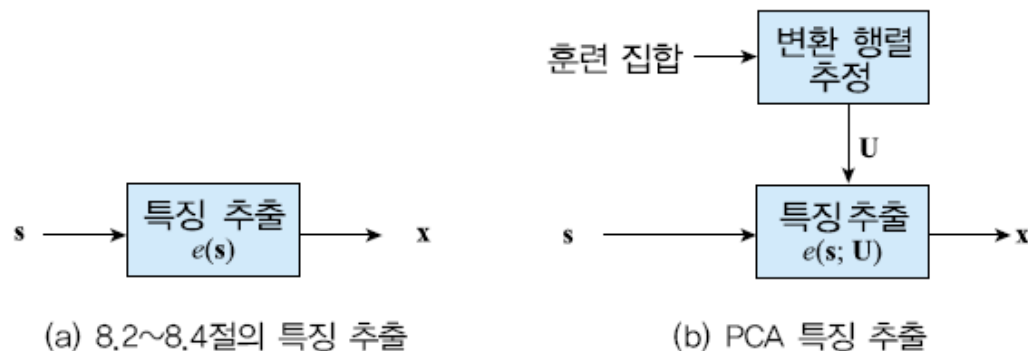


그림 8.12 특징 추출 방법의 비교

8.5.1 동기

- 차원 축소의 표현
 - D 차원 단위 벡터 \mathbf{u} 축으로의 투영

$$\hat{\mathbf{x}} = \mathbf{u}^T \mathbf{s} \quad (8.27)$$

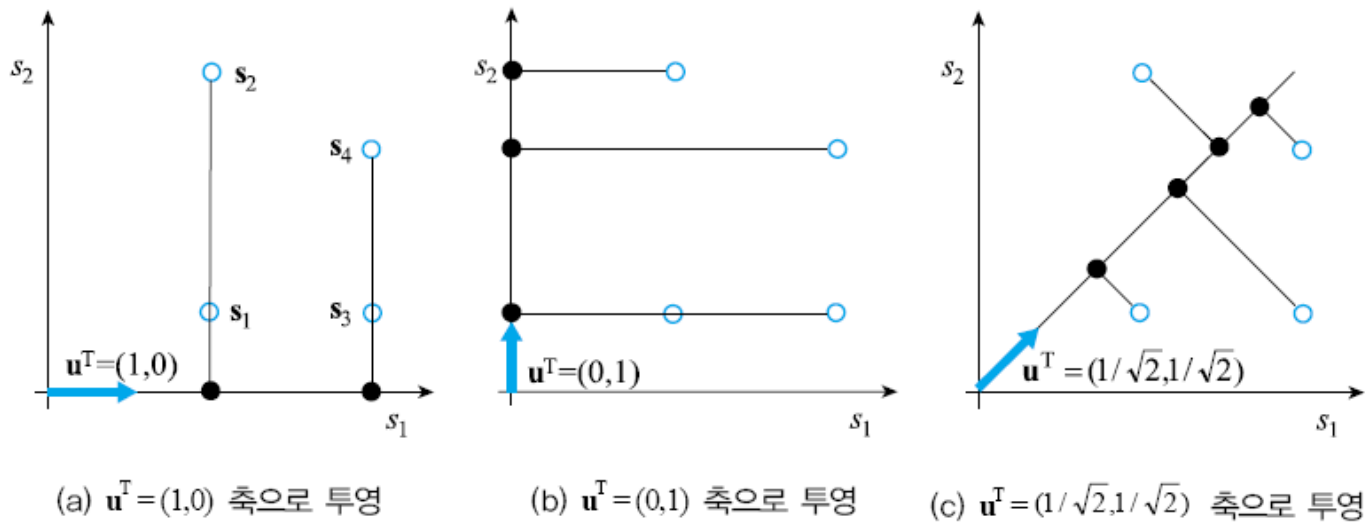


그림 8.13 투영에 의해 2 차원 공간을 1 차원 공간으로 축소

8.5.1 동기

- 정보 손실의 공식화
 - 원래 훈련 집합이 가진 정보란 무엇일까?
 - 샘플들 간의 거리, 그들 간의 상대적인 위치 등
 - 그림 8.13의 세 가지 축 중에 어느 것이 정보 손실이 가장 적은가?
 - PCA는 샘플들이 원래 공간에 ‘퍼져있는 정도를’ 변환된 공간에서 얼마나 잘 유지하느냐를 척도로 삼음
 - 이 척도는 변환된 공간에서 샘플들의 분산으로 측정함
- 이러한 아이디어에 따라 문제를 공식화 하면,
 - 변환된 샘플들의 분산을 최대화하는 축 (즉 단위 벡터 \mathbf{u})을 찾아라. (8.28)

8.5.1 동기

■ 예제 8.6 변환 공간에서의 분산

원래 샘플

$$\mathbf{s}_1 = (2,1)^T, \mathbf{s}_2 = (2,4)^T, \mathbf{s}_3 = (4,1)^T, \mathbf{s}_4 = (4,3)^T$$

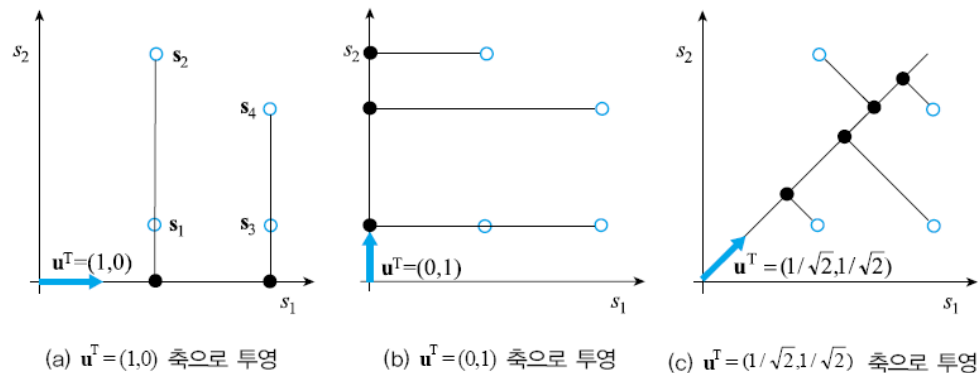


그림 8.13 투영에 의해 2 차원 공간을 1 차원 공간으로 축소

$\mathbf{u}^T = (1,0)$ 축으로 투영 변환된 샘플 분산 1.0

$$\hat{x}_1 = (1 \ 0) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2, \quad \hat{x}_2 = (1 \ 0) \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2, \quad \hat{x}_3 = (1 \ 0) \begin{pmatrix} 4 \\ 1 \end{pmatrix} = 4, \quad \hat{x}_4 = (1 \ 0) \begin{pmatrix} 4 \\ 3 \end{pmatrix} = 4$$

$\mathbf{u}^T = (1/\sqrt{2}, 1/\sqrt{2})$ 축으로 투영 변환된 샘플: 분산 1.0938

$$\hat{x}_1 = (1/\sqrt{2} \ 1/\sqrt{2}) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{3}{\sqrt{2}}$$

$$\hat{x}_2 = (1/\sqrt{2} \ 1/\sqrt{2}) \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \frac{6}{\sqrt{2}}$$

$$\hat{x}_3 = (1/\sqrt{2} \ 1/\sqrt{2}) \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \frac{5}{\sqrt{2}}$$

$$\hat{x}_4 = (1/\sqrt{2} \ 1/\sqrt{2}) \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \frac{7}{\sqrt{2}}$$

• 더 좋은 축이 있나?

8.5.2 알고리즘과 응용

- 최적의 축을 찾아 보자.

- 투영된 점의 평균과 분산

$$\hat{x}_i, 1 \leq i \leq N \text{의 평균} \quad \bar{\hat{x}} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i = \frac{1}{N} \sum_{i=1}^N \mathbf{u}^T \mathbf{s}_i = \mathbf{u}^T \left(\frac{1}{N} \sum_{i=1}^N \mathbf{s}_i \right) = \mathbf{u}^T \bar{\mathbf{s}} \quad (8.29)$$

$$\hat{x}_i, 1 \leq i \leq N \text{의 분산} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - \bar{\hat{x}})^2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{u}^T \mathbf{s}_i - \mathbf{u}^T \bar{\mathbf{s}})^2 \quad (8.30)$$

- 문제를 다시 쓰면,

- (8.30)의 분산 $\hat{\sigma}^2$ 을 최대화하는 \mathbf{u} 를 찾아라. (8.31)

- $\mathbf{u}^T \mathbf{u} = 1$ 이라는 조건을 만족하는 조건부 최적화 문제로 다시 쓰면,

- L 은 라그랑제 함수, λ 는 라그랑제 승수

- $L(\mathbf{u}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{u}^T \mathbf{s}_i - \mathbf{u}^T \bar{\mathbf{s}})^2 + \lambda(1 - \mathbf{u}^T \mathbf{u})$ 를 최대화하는 \mathbf{u} 를 찾아라. (8.32)

8.5.2 알고리즘과 응용

- 미분하고 수식 정리하면,

$$\begin{aligned}\partial L(\mathbf{u})/\partial \mathbf{u} &= \partial \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{u}^T \mathbf{s}_i - \mathbf{u}^T \bar{\mathbf{s}})^2 + \lambda(1 - \mathbf{u}^T \mathbf{u}) \right) / \partial \mathbf{u} \\ &= \frac{2}{N} \sum_{i=1}^N (\mathbf{u}^T \mathbf{s}_i - \mathbf{u}^T \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}}) - 2\lambda \mathbf{u} \\ &= 2\mathbf{u}^T \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}}) \right) - 2\lambda \mathbf{u} \\ &= 2\mathbf{u}^T \mathbf{\Sigma} - 2\lambda \mathbf{u} \\ &= 2\mathbf{\Sigma} \mathbf{u} - 2\lambda \mathbf{u}\end{aligned}$$

- 0으로 놓고 풀면,

$$\mathbf{\Sigma} \mathbf{u} = \lambda \mathbf{u} \tag{8.33}$$

- (8.33)을 해석하면,
 - 훈련 집합의 공분산 행렬 $\mathbf{\Sigma}$ 를 구하고, 그것의 고유 벡터를 구하면 그것이 바로 최대 분산을 갖는 \mathbf{u} 가 됨

8.5.2 알고리즘과 응용

■ 예제 8.7 최대 분산을 갖는 축

$X = \{(2,1)^T, (2,4)^T, (4,1)^T, (4,3)^T\}$ 훈련 집합

$$\Sigma = \begin{pmatrix} 1.000 & -0.250 \\ -0.250 & 1.688 \end{pmatrix} \text{ 공분산 행렬}$$

$\lambda_1 = 1.7688, \mathbf{u}_1^T = (-0.3092, 0.9510)$ 고유 벡터

$\lambda_2 = 0.9187, \mathbf{u}_2^T = (-0.9510, -0.3092)$

$\mathbf{u}_1^T = (-0.3092, 0.9510)$ 축으로 투영된 특징 벡터:

$$\hat{x}_1 = (-0.3092 \quad 0.9510) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0.3326$$

$$\hat{x}_2 = (-0.3092 \quad 0.9510) \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 3.1856$$

$$\hat{x}_3 = (-0.3092 \quad 0.9510) \begin{pmatrix} 4 \\ 1 \end{pmatrix} = -0.2858$$

$$\hat{x}_4 = (-0.3092 \quad 0.9510) \begin{pmatrix} 4 \\ 3 \end{pmatrix} = 1.6162$$

이들의 분산은 **1.7688**
그림 8.13과 비교해 보자.

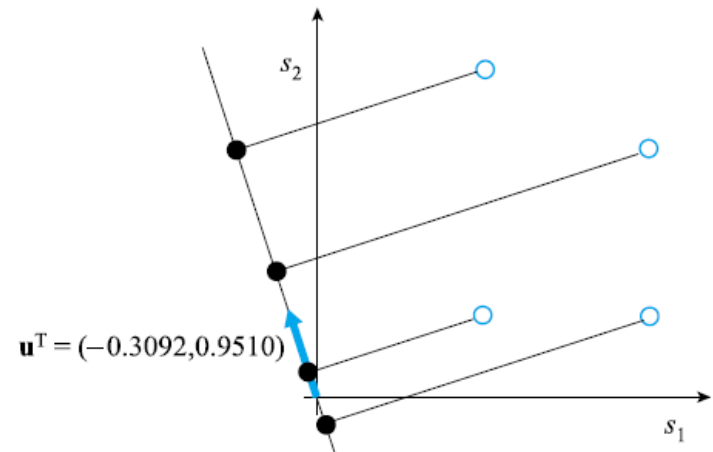


그림 8.14 PCA로 구한 최적의 축

8.5.2 알고리즘과 응용

■ 변환 행렬

- (8.33)을 풀면 D 개의 고유 벡터. 고유값이 큰 것일수록 중요도가 큼
- 따라서 D 차원을 d 차원으로 줄인다면 고유값이 큰 순으로 d 개의 고유 벡터를 취함. 이들을 주성분이라 부르고 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ 로 표기함
- 변환 행렬 \mathbf{U} 는,

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_d^T \end{pmatrix} \quad (8.34)$$

■ 실제 변환은,

$$\mathbf{x} = \mathbf{U}\mathbf{s} \quad (8.35)$$

8.5.2 알고리즘과 응용

■ 알고리즘

■ 응용

- 특징 추출 (예, 얼굴 인식)
- 차원 축소
- 데이터 압축
- 데이터 시각화

알고리즘 [8.3]

PCA에 의한 변환 행렬 구함

입력: 훈련 집합 $X = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$, 원하는 차원 d

출력: 변환 행렬 \mathbf{U} , 평균 벡터 $\bar{\mathbf{s}}$

알고리즘:

1. $\bar{\mathbf{s}} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i$ // X 의 평균 벡터
2. **for** ($i = 1$ **to** N) $\mathbf{s}'_i = \mathbf{s}_i - \bar{\mathbf{s}}$; // 평균 벡터를 빼줌
3. $\mathbf{s}'_i, 1 \leq i \leq N$ 의 공분산 행렬 Σ 를 구한다.
4. Σ 의 고유 벡터와 고유 값을 구한다.
5. 고유 값 기준으로 가장 큰 d 개의 고유 벡터를 선택한다.
이들을 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ 라 하자.
6. (8.34)로 변환 행렬 \mathbf{U} 를 만든다.
7. **return** $\mathbf{U}, \bar{\mathbf{s}}$;

알고리즘 [8.4]

PCA에 의한 특징 추출

입력: 변환 행렬 \mathbf{U} , 평균 벡터 $\bar{\mathbf{s}}$, 샘플 \mathbf{s}

출력: 특징 벡터 \mathbf{x}

알고리즘:

1. $\mathbf{s} = \mathbf{s} - \bar{\mathbf{s}}$; // 샘플에서 평균 벡터를 뺀다.
2. $\mathbf{x} = \mathbf{U}\mathbf{s}$; // (8.35)
3. **return** \mathbf{x}

8.5.3 사례 연구: 고유 얼굴

- 고유 얼굴

- 1990년대 초 Turk와 Pentland가 제안
- 얼굴 인식에서 가장 널리 쓰이는 방법 중의 하나

- 분류기 학습과 인식 단계

1. $D = n^2$ 차원을 몇 차원으로 줄일지 결정하고 그것을 d 라 한다.
2. X 와 d 를 알고리즘 [8.3]의 입력으로 주어 변환 행렬 \mathbf{U} 와 평균 벡터 $\bar{\mathbf{s}}$ 를 구한다.
3. (분류기 학습) X 의 샘플 각각을 알고리즘 [8.4]에 넣어 특징 벡터를 추출한다.
이렇게 구한 특징 벡터 집합으로 새로운 훈련 집합 X' 를 구성한다.
4. (분류기 학습) X' 로 분류기를 훈련한다. (이때 분류기는 신경망, SVM, k -NN, 트리 분류기 등 어느 것이라도 좋다.)
5. (인식) 인식을 해야 하는 새로운 얼굴 영상이 들어오면 (8.36)으로 신호 \mathbf{s} 표현으로 바꾸고 이것을 알고리즘 [8.4]에 넣어 특징 벡터 \mathbf{x} 를 추출한다. \mathbf{x} 를 분류기에 넣어 인식한다.

- 변환 행렬 \mathbf{U} 를 구성하는 고유 벡터 \mathbf{u}_i 를 고유 얼굴이라 부름

8.6 Fisher의 선형 분별

■ Fisher의 선형 분별

- 특징 추출이 아니라 분류기 설계에 해당
- 하지만 PCA와 원리가 비슷하여 8장에 배치함
- PCA와 Fisher LD는 목표가 다름
 - PCA는 정보 손실 최소화 (샘플의 부류 정보 사용 안함)
 - Fisher LD는 분별력을 최대화 (샘플의 부류 정보 사용함)

■ 원리

- 축으로의 투영 $y = \mathbf{w}^T \mathbf{x}$
- 세 개의 축 중에 어느 것이 분별력 관점에서 가장 유리한가?

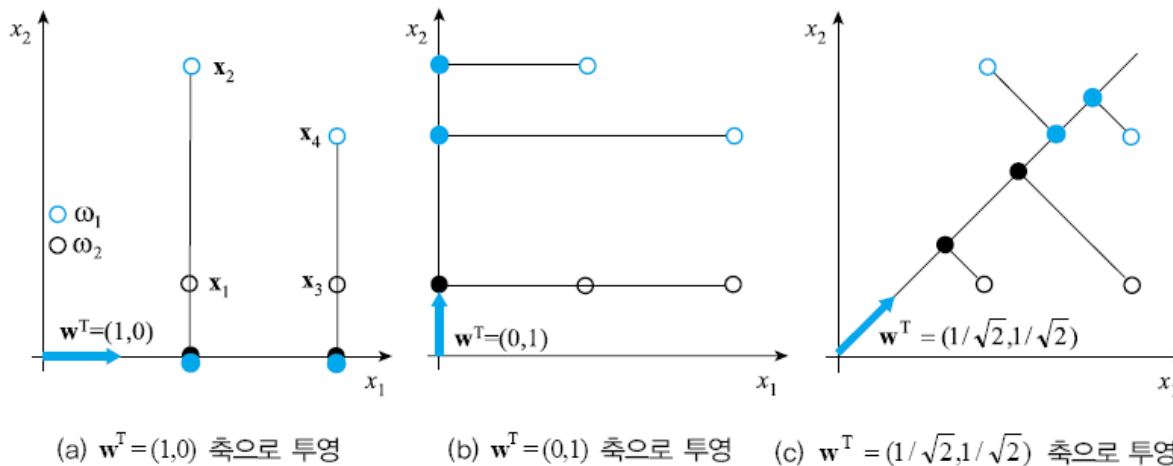


그림 8.15 2 차원 공간을 1 차원 공간으로 투영 (Fisher의 LD)

8.6 Fisher의 선형 분별

- 문제 공식화
 - 유리한 정도를 어떻게 수식화 할까?
 - 가장 유리한 축 (즉 최적의 축)을 어떻게 찾을 것인가?
- 기본 아이디어
 - “같은 부류의 샘플은 모여있고 다른 부류의 샘플은 멀리 떨어져 있을수록 유리하다.”
 - 부류간 퍼짐 between-class scatter
 - 부류내 퍼짐 within-class scatter

$$\left. \begin{aligned} \mathbf{m}_i &= \frac{1}{N_i} \sum_{\mathbf{x} \in \omega_i} \mathbf{x} \\ \bar{m}_i &= \frac{1}{N_i} \sum_{y \in \omega_i} y \\ &= \frac{1}{N_i} \sum_{\mathbf{x} \in \omega_i} \mathbf{w}^T \mathbf{x} \\ &= \mathbf{w}^T \mathbf{m}_i \end{aligned} \right\}$$

$$\text{부류간 퍼짐} = |\bar{m}_1 - \bar{m}_2| = |\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2| = |\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)|$$

$$\bar{s}_i^2 = \sum_{y \in \omega_i} (y - \bar{m}_i)^2$$

$$\text{부류내 퍼짐} = \bar{s}_1^2 + \bar{s}_2^2$$

8.6 Fisher의 선형 분별

■ 목적 함수 $J(\mathbf{w})$

$$J(\mathbf{w}) = \frac{\text{부류간 퍼짐}}{\text{부류내 퍼짐}} = \frac{|\bar{\mathbf{m}}_1 - \bar{\mathbf{m}}_2|^2}{\bar{s}_1^2 + \bar{s}_2^2} \quad (8.43)$$

□ $J(\mathbf{w})$ 를 최대화하는 \mathbf{w} 를 찾아라.

■ 분자와 분모를 다시 쓰면,

<u>분모</u>	<u>분자</u>
$\left. \begin{aligned} \bar{s}_1^2 + \bar{s}_2^2 &= \sum_{y \in \omega_1} (y - \bar{m}_1)^2 + \sum_{y \in \omega_2} (y - \bar{m}_2)^2 \\ &= \sum_{\mathbf{x} \in \omega_1} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_1)^2 + \sum_{\mathbf{x} \in \omega_2} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_2)^2 \\ &= \sum_{\mathbf{x} \in \omega_1} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_1)(\mathbf{x} - \mathbf{m}_1)^T \mathbf{w} + \sum_{\mathbf{x} \in \omega_2} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_2)(\mathbf{x} - \mathbf{m}_2)^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{S}_W \mathbf{w} \end{aligned} \right\}$	$\left. \begin{aligned} \bar{\mathbf{m}}_1 - \bar{\mathbf{m}}_2 ^2 &= (\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2)^2 \\ &= \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{S}_B \mathbf{w} \end{aligned} \right\}$
	<p>○이 때 $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$</p>

○이 때 $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$ 이고, $\mathbf{S}_i = \sum_{\mathbf{x} \in \omega_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$

8.6 Fisher의 선형 분별

- 목적 함수를 다시 쓰면,

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (8.48)$$

- $\partial J(\mathbf{w}) / \partial \mathbf{w} = 0$ 으로 두고 풀면,

$$(\mathbf{w}^T \mathbf{S}_W \mathbf{w}) \mathbf{S}_B \mathbf{w} = (\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} \quad (8.49)$$

- (8.49)를 정리하면, $(\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} = (\mathbf{w}^T \mathbf{S}_W \mathbf{w}) \mathbf{S}_B \mathbf{w}$
 $(\mathbf{w}^T \mathbf{S}_B \mathbf{w}) \mathbf{S}_W \mathbf{w} = (\mathbf{w}^T \mathbf{S}_W \mathbf{w}) \alpha_1 (\mathbf{m}_1 - \mathbf{m}_2)$

$$\mathbf{S}_W \mathbf{w} = \frac{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}{\mathbf{w}^T \mathbf{S}_B \mathbf{w}} \alpha_1 (\mathbf{m}_1 - \mathbf{m}_2)$$

$$\mathbf{S}_W \mathbf{w} = \alpha_2 \alpha_1 (\mathbf{m}_1 - \mathbf{m}_2)$$

$$\mathbf{w} = \alpha_2 \alpha_1 \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2)$$

- 결국 답은 (즉 구하고자 한 최적의 축은),

$$\mathbf{w} = \alpha \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \quad (8.50)$$

8.6 Fisher의 선형 분별

■ 예제 8.8 Fisher의 선형 분별

ω_1 샘플 (파랑): $\mathbf{x}_2 = (2,4)^T$, $\mathbf{x}_4 = (4,3)^T$

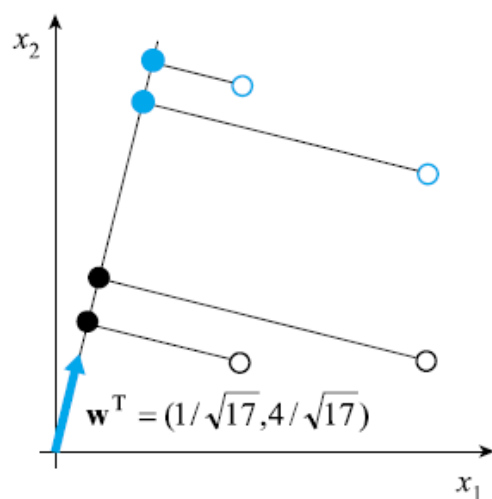
ω_2 샘플 (검정): $\mathbf{x}_1 = (2,1)^T$, $\mathbf{x}_3 = (4,1)^T$

$$\mathbf{m}_1 = (3,3.5)^T$$

$$\mathbf{m}_2 = (3,1)^T$$

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 = \left[\begin{pmatrix} -1 \\ 0.5 \end{pmatrix} (-1 \ 0.5) + \begin{pmatrix} 1 \\ -0.5 \end{pmatrix} (1 \ -0.5) \right] + \left[\begin{pmatrix} -1 \\ 0 \end{pmatrix} (-1 \ 0) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) \right] = \begin{pmatrix} 4 & -1 \\ -1 & 0.5 \end{pmatrix}$$

$$\mathbf{S}_W^{-1} = \begin{pmatrix} 0.5 & 1 \\ 1 & 4 \end{pmatrix}$$



$$\mathbf{w} = \alpha \begin{pmatrix} 0.5 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 2.5 \end{pmatrix} = \alpha' \begin{pmatrix} 1 \\ 4 \end{pmatrix}$$

$$\mathbf{w} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}^T = (0.24254, 0.97014)^T$$

이것이 최적의 축이다.
그림 8.15와 비교해 보자.

그림 8.16 Fisher의 선형 분별

Ronald Aylmer Fisher

(1890년 2월 17일 ~ 1962년 7월 29일) 영국

패턴 인식에서 Fisher가 등장하는 곳은 크게 두 군데다. 하나는 최대 우도이고 다른 하나는 Fisher의 선형 분별이다. 그가 정립한 최대 우도 이론을 역사적으로 조망한 흥미로운 논문이 있다 [Aldrich97]. 테스트용 데이터베이스로 널리 쓰이는 Iris 데이터도 그가 만들었다. 이런 이유로 그가 통계학자인 것으로 알려져 있는데 사실은 통계학뿐 아니라 유전학에도 큰 공헌을 하였다. ‘이기적 유전자’라는 *The selfish gene*



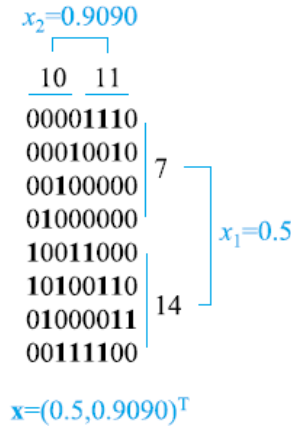
책으로 세계적인 주목을 받은 Richard Dawkins는 Fisher를 ‘다윈의 후계자중에 최고’라고 평할 정도이다. Fisher는 우생학에 *eugenics* 많은 관심을 가졌으며 Cambridge 대학 시절 우생학 그룹을 결성하여 Charles Darwin의 아들인 Horace Darwin과 같이 활동하기도 하였다. 그는 우생학을 유전학과 통계학의 접점으로 간주하였다. 그는 다윈의 자연 선택은 *natural selection* 알려진 것보다 강한 힘으로 작용한다는 연구 결과를 제시하기도 하였다. Fisher는 인도 통계 연구원을 방문하여 Mahalanobis와 교류하기도 하였으며 1957년에 Cambridge 대학을 은퇴한 후에는 호주 Adelaide 대학에서 말년을 보냈다. 그는 인종이 선천적으로 능력 차이를 지닌다는 주장을 펴기도 하였으며, 담배와 폐암의 상관 관계를 부정하는 주장을 펴기도 하였다. Fisher에 대한 보다 자세한 내용은 Adelaide 대학의 웹 [Adelaide(웹)] 또는 [Yates63]을 참고하라.

8.7 실용적 관점

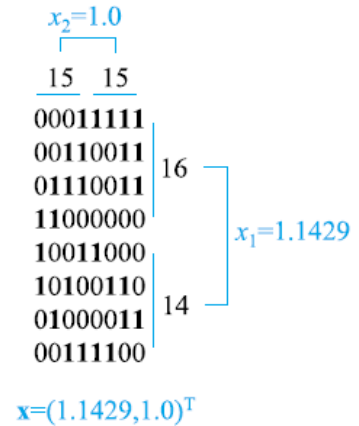
- 특징 추출은 패턴인식 과정에서 휴리스틱한 경험과 실험에 따른 시행 착오가 가장 많이 필요한 단계
 - 외부 환경에 영향을 가장 많이 받기 때문
 - 여기서 소개하는 몇가지 실용적 방법이 도움이 됨
 - 특징이 만족스러운 성능을 보이지 못하면?
 - 특징이 거리 개념을 가지지 않으면?
 - 특징마다 동적 범위가 크게 다르면?
 - 손실 특징은?

8.7.1 특징 결합

- 특징의 분별력 한계
 - 그림 8.17은 필기 숫자 예
 - 그래도 만족스럽다면 그것으로 특징 설계 완료



(a) 경향을 따르는 상황



(b) 경향을 어기는 상황

그림 8.17 특징의 성질 (숫자 부류 6)

- 특징이 만족스럽지 않다면,
 - 버리고 다른 특징을 채택
 - 또는 기존 특징에 새로운 특징을 추가하는 특징 결합
- 특징이 가지는 정보
 - 전역 정보
 - 예) 검은 화소 비율
 - 지역 정보
 - 예) 프로파일

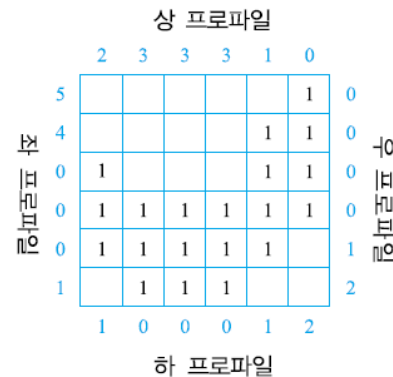


그림 8.7 프로파일 특징

8.7.2 특징 전처리

- 거리 개념이 없는 특징의 변환

- 예) 혈액형을 나타내는 특징 $x \in \{A, B, O, AB\}$
- 거리 개념이 없는 특징 x_i 가 n 개의 값을 갖는다면 x_i 를 $x_{i1}, x_{i2}, \dots, x_{in}$ 으로 확장
- $x_{i1}, x_{i2}, \dots, x_{in}$ 중 하나만 1을 가지고 나머지는 0

- 특징 값의 정규화

- 선형 변환

$$\tilde{x}_i = low_i + \frac{high_i - low_i}{max_i - min_i} (x_i - min_i) \quad (8.51)$$

- 통계에 의한 변환 (평균은 0, 표준 편차는 1을 가지도록 정규화)

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{\sigma_i} \quad (8.52)$$

8.7.2 특징 전처리

■ 예제 8.9 특징 정규화

- 사람을 키 (m 단위)와 몸무게 (kg 단위)의 두 개 특징으로 표현

$$\mathbf{a} = (1.60, 70.0)^T, \mathbf{b} = (1.65, 65.5)^T, \mathbf{c} = (1.95, 71.0)^T, \mathbf{d} = (1.68, 72.0)^T$$

- 거리 계산에 따르면 \mathbf{a} 는 \mathbf{b} 보다 \mathbf{c} 에 가깝다.
- 몸무게의 동적 범위가 커서 거리 계산을 주도하기 때문

- (8.51)의 정규화 식을 유도하면,

$$\tilde{x}_1 = 0 + \frac{1-0}{1.95-1.60}(x_1 - 1.60) = \frac{1}{0.35}(x_1 - 1.60)$$

$$\tilde{x}_2 = 0 + \frac{1-0}{72.0-65.5}(x_2 - 65.5) = \frac{1}{6.5}(x_2 - 65.5)$$

- 정규화하고 거리를 계산해 보면,

$$\mathbf{a}' = (0, 0.692)^T, \mathbf{b}' = (0.143, 0)^T, \mathbf{c}' = (1, 0.846)^T, \mathbf{d}' = (0.229, 1)^T$$

$$\text{dist}(\mathbf{a}', \mathbf{b}') = \sqrt{(0 - 0.143)^2 + (0.692 - 0)^2} = 0.707$$

$$\text{dist}(\mathbf{a}', \mathbf{c}') = \sqrt{(0 - 1)^2 + (0.692 - 0.846)^2} = 1.012$$

8.7.2 특징 전처리

- 손실 특징의 보충
 - 실제 의료나 고객 관리 등의 응용에서 발생하는 데이터가 손실 특징을 많이 가짐
 - 예) 어떤 학생이 월 수입 미기재
 - 방법
 - 훈련 집합이 충분히 큰 경우 손실 특징을 가진 샘플 제거
 - 또는 다른 샘플로 평균을 구하여 손실 샘플에 채움
 - 또다른 여러 방법들