08. 특징 추출

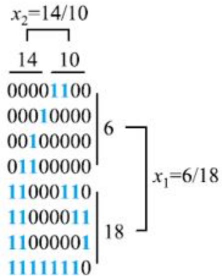
• 특징 추출의 예

6 → ·크기 정규화 ·이진화 →

특징 추출

→ 방법 1: 화소 각각을 특징으로 삼음
64차원 특징 벡터
x= (0, 0, 0, 0, 1, 1, ..., 1, 1, 0)^T

→ 방법 2: 가로 이등분과 세로 이등분하여 검은 화소의 개수 비율을 특징으로 삼음



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

→ 또 다른 방법으로는?

• 특징의 우수성 기준

① 분별력 (discriminatory) : 좋은 특징은 서로 다른 부류를 잘 분별해 주어야 함

② 차원 (dimensionality) : 특징 벡터의 차원이 낮을수록 계산 효율이 좋고 차원의 저주에서 멀어진다.

8.1.1 실제 세계의 다양성

• 특징 생성

– 외부의 물리적 패턴을 특징 벡터라는 수학적 표현으로 변환

실제 세계 —추상화→ 수학세계

패턴 → $\boxed{\text{특징 생성}}$ $\xrightarrow{\text{특징벡터 } x}$ $\boxed{\text{분류}}$ → 부류

- 특징 생성 과정은 매우 다양
  - 특징 추출은 외부 환경에 맞게 설계해야 하기 때문

8.1.2 특징 추출과 특징 선택

- 센싱으로 얻은 신호의 다양성
  - 영상, 시간성 신호, 측정 벡터

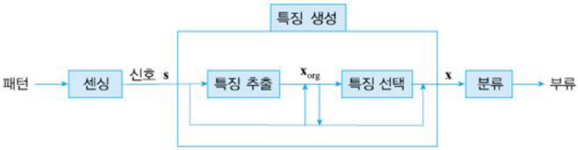


그림 8.2 특징 생성의 절차

$$\left.\begin{array}{ll} \text{특징추출} & \mathbf{x}_{org} = e(\mathbf{s}) \\ \text{특징선택} & \mathbf{x} = s(\mathbf{x}_{org}) \end{array}\right\}$$

- 다양한 상황

$$x=s,\quad x=e(s),\quad x=s(s),\quad x=s(e(s))$$

8.2 영역에서의 특징 추출

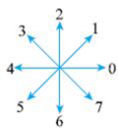
- 영역의 표현

	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)-2-2-7-0-0-1-2-1-6-6-6-5-5-4-4-3

(c) 체인 코드 표현

그림 8.3 영역의 표현

· 모멘트와 중심 모멘트

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

· 여러 가지 특징들

면적 $a = m_{00}$

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

행 분산 $V_{rr} = \frac{\overline{m}_{20}}{a}$

열 분산 $V_{cc} = \frac{\overline{m}_{02}}{a}$

혼합 분산 $V_{rc} = \frac{\overline{m}_{11}}{a}$

둘레 $p = n_{even} + n_{odd}\sqrt{2}$

이 때 n_{even} = 짝수를 갖는 체인의 개수

n_{odd} = 홀수를 갖는 체인의 개수

둥근정도 $r = \frac{4\pi a}{p^2}$ ← area ← perimeter

· 불변 특성

- 이동 불변 (translation invariant)

특징 a는 영역 R이 다른 위치로 이동해도 값이 변하지 X

중점을 나타내는 특징 $\overline{x}$와 $\overline{y}$는 이동 불변이지 않다.

- 크기 불변 (scale invariant)

영역의 크기가 중점을 기준으로 변한다면 중점은 그대로 유지된다.

특징 a는 크기 불변 X

- 회전 불변 (rotation invariant)

$\overline{x}$와 $\overline{y}$는 자신을 중심으로 회전하면 회전 불변이지만, 다른 점을 중심으로 회전하면 회전 불변 X

예제 8.1 모양 특징의 추출

	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

1번 영역의 특징

면적 $a=20$

중점 $(\bar{x},\bar{y}) = \left(\frac{\sum_{(x,y)\in R} x}{20} , \frac{\sum_{(x,y)\in R} x}{20} \right)$

$= (x$의 합$/20,\ y$의 합$/20) = (2.7, 3.05)$

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

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

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

둘레 $P = 10+6\sqrt{2} = 18.485$

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

	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

→ 원이 디지털 공간에 근사화되어 있기 때문에 발생하는 현상

→ 원이 커지면 오류 적어짐

8.2.2 투영과 프로파일 특징

- 투영(projection) 특징

  행 투영 h_i = 행 i에 있는 1의 개수, $0 \le i \le N-1$

  열 투영 v_j = 열 j에 있는 1의 개수, $0 \le j \le M-1$

  → $N+M$ 차원의 특징 벡터를 얻게 됨

◎ 몇개 건너 투영하거나, 이웃한 투영 몇 개의 평균을 구하는 방법으로 특징의 개수 줄일 수 있음. 평균을 사용하면 잡음에 둔감해지는 효과

예제 8.2

	0	1	2	3	4	5	행 투영
0						1	1
1					1	1	2
2	1				1	1	3
3	1	1	1	1	1	1	6
4	1	1	1	1	1		5
5		1	1	1			3
열 투영	3	3	3	3	4	4	

그림 8.6 투영 특징

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

두개의 투영 평균하면, $x = (3,3,4,1.5,4.5,4)^T$

• 프로파일 특징

- $2(N+M)$ 차원의 특징 벡터를 얻게 됨

상 프로파일 t_j = 위에서 열 j를 바라보았을 때 처음 1까지의 거리, $0 \le j \le M-1$

우 프로파일 r_i = 오른쪽에서 행 i를 바라보았을 때 처음 1까지의 거리, $0 \le i \le N-1$

하 프로파일 b_j = 아래에서 열 j를 바라보았을 때 처음 1까지의 거리, $0 \le j \le M-1$

좌 프로파일 l_i = 왼쪽에서 행 i를 바라보았을 때 처음 1까지의 거리, $0 \le i \le N-1$

예제 8.3 프로파일 특징의 추출

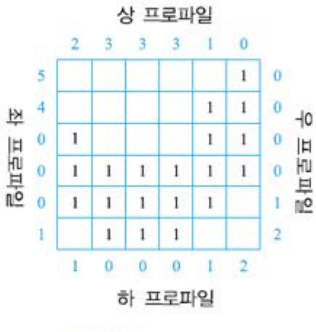


그림 8.7 프로파일 특징

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

이웃한 3개 평균? $x = (2.66, 1.33, 0, 1, 1, 0.33, 0.33, 3)^T$

8.3 변환을 이용한 특징

- 파형(waveform) 신호에서의 특징 추출
  예) 지진파, 기계 진동파, 수중파, 음파, 주식 곡선 등

- 파형에서 어떻게 특징을 추출할 것인가?
  → 파형은 기저 함수의 선형 결합으로 표현 가능 : 선형 결합의 계수를 특징으로 함. 계수를 어떻게 구할 것인가?

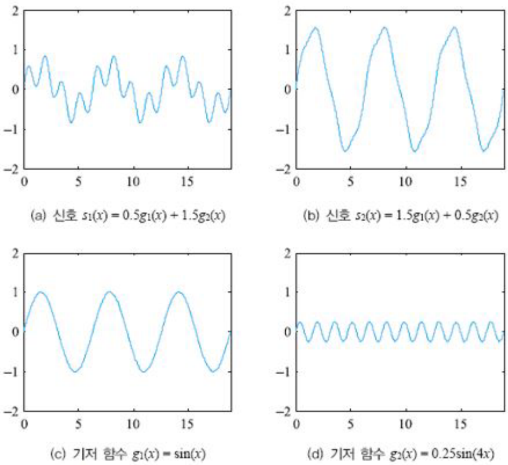


그림 8.8 입력 신호는 기저 함수의 선형 결합으로 표현할 수 있음

$s_1(x) = 0.5g_1(x) + 1.5g_2(x)$
$s_2(x) = 1.5g_1(x) + 0.5g_2(x)$

신호 s_1의 특징벡터 $= (0.5, 1.5)^T$
신호 s_2의 특징벡터 $= (1.5, 0.5)^T$

⇒ 기저함수의 계수는 어떻게 구할건데?
이산 퓨리에 변환

8.3.1 퓨리에 변환

→ 임의의 입력 신호를 다양한 주파수를 갖는 주기함수들의 합으로 분해하여 표현

- 이산 퓨리에 변환

신호를 $S = (S(0), S(1), \cdots, S(n-1))^T$로 표현할 때,
퓨리에 변환 $f(u)$는 시간 공간을 주파수 공간으로 바꾸어 준다.

$$f(u) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} S(i) \exp\left(-j \frac{2\pi u i}{n}\right), \quad u = 0, \cdots, n-1$$

예제 8.4 이산 퓨리에 변환

$S=(1.7, 0.6, 1.5, 1.2)^T$

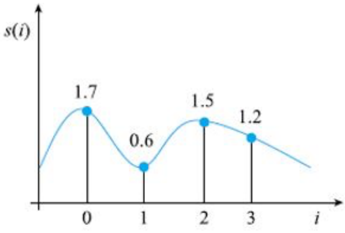


그림 8.9 신호 샘플링

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

이처럼 어떤 입력 신호든 간에 다 sin, cos 주기함수들의 합으로 분해가능

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

$$f(1)=0.5(1.7e^0+0.6e^{-j\pi/2}+1.5e^{-j2\pi/2}+1.2e^{-j3\pi/2})$$
$$=0.5(1.7+0.6(\cos(-\pi/2)+j\sin(-\pi/2))+1.5(\cos(-2\pi/2)$$
$$+j\sin(-2\pi/2))+1.2(\cos(-3\pi/2)+j\sin(-3\pi/2)))$$
$$=0.5((1.7+0.6\cos(-\pi/2)+1.5\cos(-2\pi/2)+1.2\cos(-3\pi/2))$$
$$+j(0.6\sin(-\pi/2))+1.5\sin(-2\pi/2))+1.2\sin(-3\pi/2)))=0.1+j0.3$$

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

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

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

- 퓨리에 특징

– 파워 스펙트럼의 값을 특징으로 취함

$$p(u)=\sqrt{\mathrm{real}(f(u))^2+\mathrm{imag}(f(u))^2},\quad u=0,\cdots,n-1$$

예제 8.5 퓨리에 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 4차원 특징 벡터를 얻게 된다.

$$x = (2.5, 0.316, 0.7, 0.316)^T$$

- 퓨리에 특징 알고리즘
  - d개의 특징을 추출하고자 하면 $P(0), P(1), \cdots, P(d-1)$을 취한다.
  - 왜냐하면 u가 커질수록 $P(u)$는 급속도로 작아지므로

알고리즘 [8.1] **퓨리에 특징 추출**

입력: 신호 $\mathbf{s}$, 필요한 특징 개수 d

출력: d 차원의 특징 벡터 $\mathbf{x}$

알고리즘:

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

- 2차 퓨리에 변환

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

파워 스펙트럼 $P(u,v) = \sqrt{\mathrm{real}(f(u,v))^2 + \mathrm{imag}(f(u,v))^2} \qquad (8.22)$

· 퓨리에 기술자

- 2차원 퓨리에 변환은 n×n영상에 적용할 수 있다.
- 영역 경계 상의 점을 복소수로 표현한 뒤에 2차원 퓨리에 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)

$$d(i) = x_i + jy_i, \quad 0 \le i \le n-1$$

ex) $d(4) = 2 + j3$ 처럼 표현

퓨리에 기술자 $f(u) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} d(i) \exp\left(-j\frac{2\pi u i}{n}\right), \; u = 0, \ldots, n-1$

8.4 시계열 신호에서의 특징 추출

시계열 형태 (지진파, 음성파, 주식 시세) → 동적인 성질을 가지므로 정적인 영상과 다른 추출 방식 필요 : ① 주로 퓨리에 변환 이용

② 보통 신호를 frame 단위로 나눔

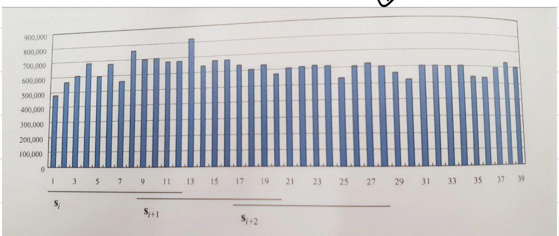


그림 8.11 시계열 신호의 프레임 단위 분할 (프레임의 크기 k=12인 경우)

각각의 프레임에서 특징 벡터 추출

시계열 신호에서 추출한 특징벡터의 열 $(x_1, x_2, \cdots, x_n)^T$

i번째 frame의 신호를 S_i로 표기하자. 특징 추출을 위한 아주 간단한 방법은 S_i를 알고리즘 [8.1]에 입력하여 퓨리에 특징 벡터 x_i를 구하는 것이다. 특징 벡터는 (8.25)와 같이 표현할 수 있다.

$P_i(j)$는 i번째 프레임에서 얻은 파워 스펙트럼의 j번째 값이다.

$$x_i = (P_i(0), P_i(1), \cdots, P_i(d-1))^T$$

퓨리에 특징보다 분별력에서 우수하여 보다 널리 쓰이는 cepstrum 특징이 있다.

알고리즘 [8.2] 캡스트럼 특징 추출

입력: 신호 $\mathbf{s} = (s(0), s(1), \cdots, s(n-1))$, 확장 프레임 길이 m, 필요한 특징 개수 d

출력: d 차원의 특징 벡터 $\mathbf{x}$

알고리즘:

1. $\mathbf{s}$의 뒤에 $m-n$개의 0을 붙여 $\mathbf{s}' = (s(0), s(1), \cdots, s(n-1), 0, 0, \cdots, 0)$을 구한다.
2. $\mathbf{s}'$에 대해 (8.19)로 이산 퓨리에 변환을 수행하여 $f(u)$를 구한다.
3. $f(u)$의 파워 스펙트럼에 로그를 취하고 그 결과에 대해 퓨리에 역 변환을 취한다.
4. 3에서 구한 결과를 특징 벡터 $\mathbf{x}$로 취한다.

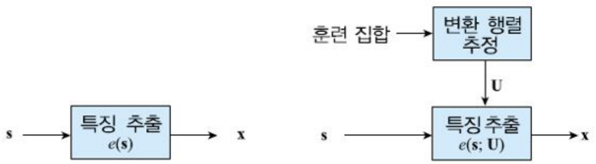
8.5 주성분 분석

- 훈련 집합을 이용하여 매개 변수를 추정하고 그것을 이용하여 특징을 추출함
- 정보 손실을 최소화하는 조건에서 차원 축소
- karhunen-Loeve(KL) 변환 또는 Hotelling 변환이라고 부름

8.5.1 동기

- 주성분 분석의 동기
  - U는 '정보 손실을 최소화하며' 신호 s를 보다 낮은 차원의 특징벡터 x로 변환 ($d < D$)
  - 변환 행렬 U는 $d \times D$ 행렬 (신호 s는 D차원, 특징벡터 x는 d차원)
  - 두가지 문제
    ① 차원 축소를 어떻게 표현할 것인가?
    ② 정보 손실을 어떻게 수량화할 것인가?

$$x = e(s; U) = Us$$

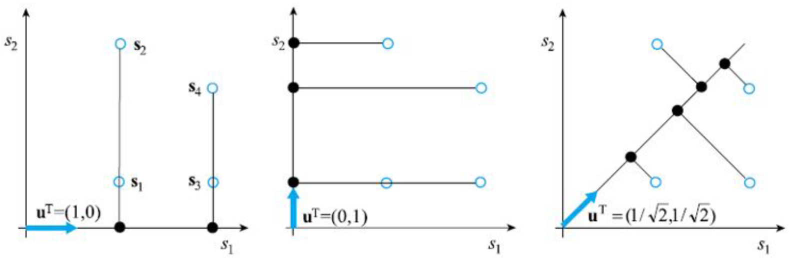


(a) 8.2~8.4절의 특징 추출 (b) PCA 특징 추출

그림 8.12 특징 추출 방법의 비교

- 차원 축소의 표현
  - D차원 단위벡터 u축으로의 투영

$$\hat{x} = u^T s$$



(a) $\mathbf{u}^T=(1,0)$ 축으로 투영 (b) $\mathbf{u}^T=(0,1)$ 축으로 투영 (c) $\mathbf{u}^T=(1/\sqrt{2},1/\sqrt{2})$ 축으로 투영

그림 8.13 투영에 의해 2 차원 공간을 1 차원 공간으로 축소

• 정보 손실의 공식화

- 원래 훈련 집합이 가진 정보란 무엇일까?
  샘플들 간의 거리, 그들 간의 상대적인 위치
- PCA는 샘플들이 원래 공간에 '퍼져있는 정도를' 변환된 공간에서 얼마나 잘 유지하느냐를 척도로 삼음
  → 이 척도는 변환된 공간에서 샘플들의 분산으로 측정

변환된 샘플들의 분산을 최대화하는 축(즉 단위 벡터 u)을 찾아라.

예제 8.6 변환 공간에서의 분산

원래 샘플 $s_1=(2,1)^T, s_2=(2,4)^T, s_3=(4,1)^T, s_4=(4,3)^T$

$u^T=(1,0)$축으로 투영된 샘플

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

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

분산 : 1.0

$u^T=(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$축으로 투영된 샘플

$\hat{x}_1=(\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}})\begin{pmatrix}2\\1\end{pmatrix}=\frac{3}{\sqrt{2}} \quad \hat{x}_3=(\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}})\begin{pmatrix}4\\1\end{pmatrix}=\frac{5}{\sqrt{2}}$

$\hat{x}_2=(\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}})\begin{pmatrix}2\\4\end{pmatrix}=\frac{6}{\sqrt{2}} \quad \hat{x}_4=(\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}})\begin{pmatrix}4\\3\end{pmatrix}=\frac{7}{\sqrt{2}}$

분산 : 1.0938 더 좋은 축이 있나?

8.5.2 알고리즘과 응용

- 투영된 점의 평균과 분산

$\hat{x}_i,\ 1 \le i \le N$의 평균: $\bar{\hat{x}} = \frac{1}{N}\sum_{i=1}^{N}\hat{x}_i = \frac{1}{N}\sum_{i=1}^{N}u^T s_i = u^T\left(\frac{1}{N}\sum_{i=1}^{N}s_i\right) = u^T\bar{s}$ (8.29)

$\hat{x}_i,\ 1 \le i \le N$의 분산: $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - \bar{\hat{x}})^2 = \frac{1}{N}\sum_{i=1}^{N}(u^T s_i - u^T\bar{s})^2$ (8.30)

즉, (8.30)의 분산 $\hat{\sigma}^2$을 최대화하는 u를 찾아라.

u가 단위벡터라는 조건을 이용하면 $u^T u = 1$이라는 조건을 만들 수 있다.

조건부 최적화 문제로 다시 쓰면,

- L은 라그랑제 함수, λ는 라그랑제 승수

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

$$\partial L(u)/\partial u = \partial\left(\frac{1}{N}\sum_{i=1}^{N}(u^T s_i - u^T\bar{s})^2 + \lambda(1 - u^T u)\right)/\partial u$$

$$= \frac{2}{N}\sum_{i=1}^{N}(u^T s_i - u^T\bar{s})(s_i - \bar{s}) - 2\lambda u$$

$$= 2u^T\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N}(s_i - \bar{s})^2}_{\text{공분산행렬 (대칭임!)}}\right) - 2\lambda u$$

$$= 2u^T\Sigma - 2\lambda u$$

$$= 2\Sigma u - 2\lambda u = 0$$

$\Sigma u = \lambda u$ → u는 공분산 행렬 Σ의 고유 벡터, λ는 고유값

고유벡터: 선형 변환 A에 의한 변환 결과가 자기 자신의 상수배가 되는 0이 아닌 벡터

행공간에 수직인 벡터

예제 8.7 최대 분산을 갖는 축

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

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

고유값 $\lambda_1 = 1.7688$, $u_1^T = (-0.3092, 0.9510)$

$\lambda_2 = 0.9187$, $u_2^T = (-0.9510, -0.3092)$

$u_1^T = (-0.3092, 0.9510)$ 으로 투영된 특징벡터

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

$\hat{x}_2 = 3.1856$ $\hat{x}_3 = -0.2858$, $\hat{x}_4 = 1.6162$

평균: $(-0.3092, 0.9510)\begin{pmatrix} 3 \\ \frac{9}{4} \end{pmatrix} = 1.2122$

분산: 1.7688

지금까지 하나의 축으로 투영하는 상황설명. 실제에서는 D차원을 d차원으로 줄이는 것이 목적이므로 d개의 축을 찾아야 한다.

고유벡터의 성질에 따라 이들은 모두 수직(orthogonal)이다.

즉, $u_i^T u_j = 0$, $i \neq j$ 이다. 또한, 고유 벡터는 고유 값을 갖는데 고유값이 클수록 중요도가 크다. 따라서 고유값에 따라 정렬을 하고, 상위 d개의 고유 벡터를 선정하면 된다.

이들을 $u_1, u_2, \cdots, u_d$로 표기하자. 이들 각각을 주성분 (principal component)라 부른다. U의 한 행을 차지하는 u_i^T는 크기가 D인 행 벡터이다. 따라서 U는 d×D 행렬이다.

$$U = \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_d^T \end{pmatrix}$$

이제 PCA로 구한 변환행렬 U를 이용하여 신호 s를 특징벡터 x로 변환하는 식이다.

알고리즘 [8.3] PCA에 의한 변환 행렬 구함

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

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

알고리즘:

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

좌표계의 원점이 평균 벡터 $\bar{s}$와 일치하도록 해주는 전처리 과정

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** $\mathbf{x}$

· PCA의 응용

① 특징 추출(얼굴인식) → 영역분할의 어려움으로 성능 한계에 직면한 것이 사실

② 특징 벡터의 차원 축소

PCA : 훈련집합 샘플들은 부류 정보 X → 원래 특징 벡터가 가진 정보의 손실을 최소화하는 변환행렬을 찾음

LDA : 훈련집합 부류정보 O → 변환공간에서 서로 다른 부류가 가장 잘 분별될 수 있는 변환행렬을 찾음

③ 데이터 압축

④ 시각화

8.5.3 사례 연구 : 고유 얼굴

얼굴인식은 눈, 코, 입 등의 정확한 영역 분할이 어려워 정확한 특징 값을 얻는다는 보장이 없음.

→ 1990년 대 초 Turk와 Pentland가 색다른 방법 제안 :

고유 얼굴(eigenface)

2차원 영상 표현을 1차원 신호 표현으로 바꾼다!

$$S=(s_1, s_2, \cdots, s_D)^T = (I[1][1], I[1][2], \cdots, I[1][n], I[2,1], \cdots I[n][n-1], I[n][n])^T \quad (8.36)$$

(얼굴 영상 집합 $Y=\{I_1, I_2, \cdots, I_N\}$. I는 $n \times n$ 크기의 영상, $I[r][c]$, $1 \le r, c \le n$으로 표기

• 분류기 학습과 인식단계

1. $D=n^2$ 차원을 몇 차원으로 줄일지 결정(d)

2. X와 d를 알고리즘 [8.3]의 입력으로 주어 변환 행렬 U와 평균 벡터 s를 구한다.

3. (분류기 학습) X의 샘플 각각을 알고리즘 [8.4]에 넣어 특징 벡터 추출

4. (분류기 학습) X'로 분류기 훈련(이때 분류기는 신경망, SVM, k-NN, 트리 분류기 등 어느 것이라도 좋다)

5. (인식) 인식을 해야하는 새로운 영상이 들어오면 (8.36)으로 신호 s 표현을 바꾸고 이를 알고리즘 [8.4]에 넣어 특징벡터 x를 추출. x를 분류기에 넣어 인식한다.

· 변환 행렬 U를 구성하는 고유벡터 u_i를 고유 얼굴이라 부름.

8.6 Fisher의 선형 분별

• 특징 추출이 아닌 분류기 설계에 해당

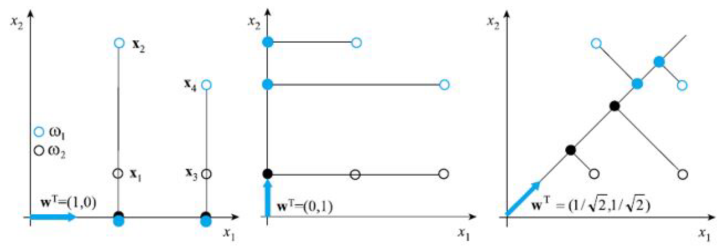
• but, PCA와 원리가 비슷하여 이곳에 배치

• PCA와 Fisher LD는 목표가 다름

- PCA는 정보 손실 최소화 (샘플의 부류 정보 사용 X)
- Fisher LD는 분별력을 최대화 (샘플의 부류 정보 사용 O)

• 원리

- 축으로의 투영 $y=w^Tx$
- 세 개의 축 중에 어느것이 분별력 관점에서 가장 유리한가?



(a) $\mathbf{w}^T=(1,0)$ 축으로 투영 (b) $\mathbf{w}^T=(0,1)$ 축으로 투영 (c) $\mathbf{w}^T=(1/\sqrt{2},1/\sqrt{2})$ 축으로 투영

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

• 문제 공식화

- 유리한 정도를 어떻게 수식화할까?
- 가장 유리한 축을 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)

N_i는 w_i에 속하는 샘플의 개수이다.

$$m_i = \frac{1}{N_i}\sum_{x\in w_i} x \quad \text{(원래 공간에서의 평균 점)} \qquad (8.38)$$

$$\overline{m}_i = \frac{1}{N_i}\sum_{y\in w_i} y \quad \text{(투영 후의 평균 점)} \qquad (8.39)$$

$$= \frac{1}{N_i}\sum_{x\in w_i} w^T x$$

$$= w^T m_i$$

$$\text{부류간 퍼짐} = |\overline{m}_1 - \overline{m}_2| = |w^T m_1 - w^T m_2| = |w^T(m_1 - m_2)| \qquad (8.40)$$

$$\overline{s}_i^{\,2} = \sum_{y\in w_i}(y - \overline{m}_i)^2 \qquad (8.41)$$

$$\text{부류내 퍼짐} = \overline{s}_1^{\,2} + \overline{s}_2^{\,2} \qquad (8.42)$$

• 목적 함수 $J(w)$

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

$J(w)$를 최대화하는 w를 찾아라.

• 분자와 분모를 다시 쓰면,

분모 $\tilde{s}_1^2 + \tilde{s}_2^2 = \sum_{y \in w_1} (y - \tilde{m}_1)^2 + \sum_{y \in w_2} (y - \tilde{m}_2)^2$

$= \sum_{x \in w_1} (w^T x - w^T m_1)^2 + \sum_{x \in w_2} (w^T x - w^T m_2)^2$

$= \sum_{x \in w_1} w^T (x - m_1)(x - m_1)^T w + \sum_{x \in w_2} w^T (x - m_2)(x - m_2)^T w$

$= w^T S_w w \qquad (8.44)$

이 때 $S_w = S_1 + S_2$이고, $S_i = \sum_{x \in w_i} (x - m_i)(x - m_i)^T$

부류내 퍼짐 행렬

분자 $|\tilde{m}_1 - \tilde{m}_2|^2 = (w^T m_1 - w^T m_2)^2$

$= w^T (m_1 - m_2)(m_1 - m_2)^T w$

$= w^T S_B w \qquad (8.45)$

이 때 $S_B = (m_1 - m_2)(m_1 - m_2)^T$

부류간 퍼짐 행렬

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

$\frac{\partial J(w)}{\partial w} = 0$으로 두고 풀면,

$(w^T S_w w) S_B w = (w^T S_B w) S_w w$

$$(w^T S_B w) S_w w = (w^T S_w w) S_B w$$

$$(w^T S_B w) S_w w = (w^T S_w w) \alpha_1 (m_1 - m_2)$$

$$S_w w = \frac{w^T S_w w}{w^T S_B w} \alpha_1 (m_1 - m_2)$$

$$S_w w = \alpha_2 \alpha_1 (m_1 - m_2)$$

$$w = \alpha_2 \alpha_1 w^{-1} (m_1 - m_2)$$

$$w = \alpha S_w^{-1} (m_1 - m_2)$$

예제 8.8

w_1 샘플(파랑): $x_2 = (2,4)^T$, $x_4 = (4,3)^T$

w_2 샘플(검정): $x_1 = (2,1)^T$, $x_3 = (4,1)^T$

$m_1 = (3, 3.5)^T$ $m_2 = (3,1)^T$

$$S_w = S_1 + S_2 = \left[\begin{pmatrix} -1 \\ 0.5 \end{pmatrix} (-1 \ 0.5) + \begin{pmatrix} 1 \\ -0.5 \end{pmatrix} (1 \ -0.5) \right] +$$

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

$$S_w^{-1} = \begin{pmatrix} 0.5 & 1 \\ 1 & 4 \end{pmatrix}$$

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

W의 크기 무시 가능하므로,

$$w = \left(\frac{1}{\sqrt{17}} \quad \frac{4}{\sqrt{17}}\right)^T = (0.24254,\ 0.97014)^T$$

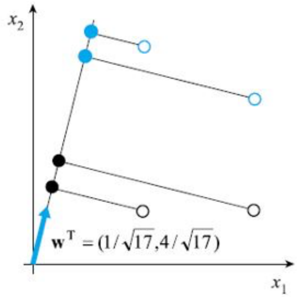


그림 8.16 Fisher의 선형 분별

8.7 실용적 관점

8.7.1 특징 결합

- 특징의 분별력 한계
  - 그림 8.17은 필기 숫자의 예
  - 그래도 만족스럽다면 그것으로 특징설계 완료

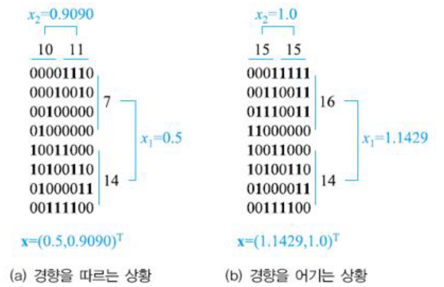


그림 8.17 특징의 성질 (숫자 부류 6)

- 특징이 만족스럽지 않다면,
  - 버리고 다른 특징을 채택
  - 또는 기존 특징에 새로운 특징을 추가하는 특징 결합
- 특징이 가지는 정보
  - 전역 정보

    ex) 검은 화소 비율
  - 지역 정보

    ex) 프로파일

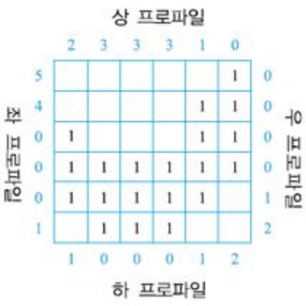


그림 8.7 프로파일 특징

8.7.2 특징 전처리

• 거리 개념이 없는 특징의 변환

- 예) 혈액형, 직업, 성씨 등
- 거리 개념이 없는 특징 x_i가 n개의 값을 갖는다면 x_i를 $x_{i1}, x_{i2}, \cdots x_{in}$으로 확장
- $x_{i1}, x_{i2}, \cdots, x_{in}$ 중 하나만 1을 가지고 나머지는 0

• 특징 값의 정규화

- 선형 변환

$$\tilde{x}_i = low_i + \frac{high_i - low_i}{\underbrace{max_i - min_i}_{x_i\text{의 동적범위}}}(x_i - min_i)$$

새로운 동적 범위

• 통계에 의한 변환 (평균은 0, 표준편차는 1을 가지도록 정규화)

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{\sigma_i}$$

• 손실 특징의 보충

데이터 손실이 있을 때,

① 훈련집합이 충분히 큰 경우 손실 특징 가진 샘플 제거

② 또는 다른 샘플로 평균을 구하여 손실 샘플에 채움

⋮