

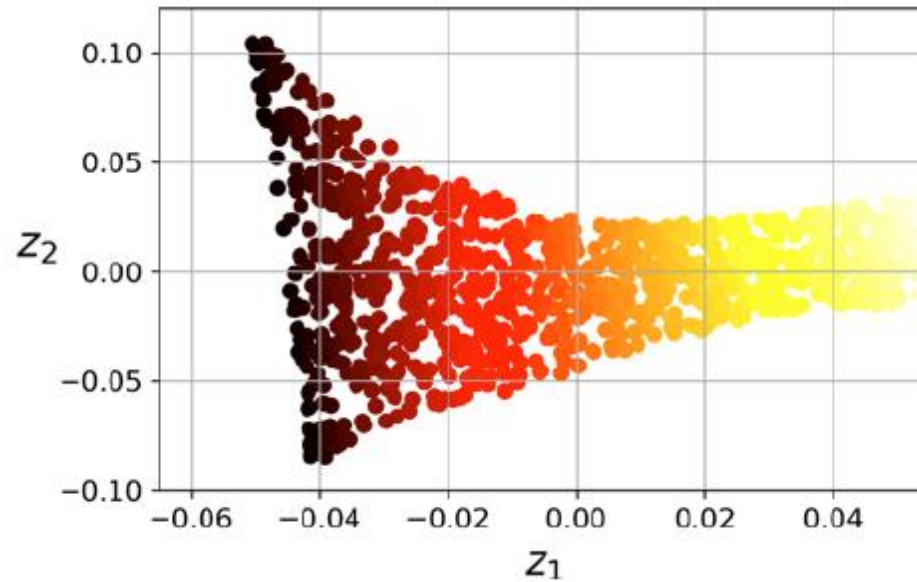
차원 축소

8-5. LLE

# 지역 선형 임베딩(LLE)

- Local Linear Embedding
- 투영에 의존하지 않는 매니폴드 학습
- 각 훈련 샘플이 가장 가까운 이웃에 얼마나 선형적으로 연관되어 있는지 측정
  - > 국부적인 관계가 가장 잘 보존되는 훈련 세트의 저차원 표현 찾기
- 잡음이 너무 많지 않은 경우 꼬인 매니폴드를 펼치는 데 잘 작동

# 지역 선형 임베딩(LLE)



- 스위스 롤이 완전히 펼쳐져 있으나, 오른쪽은 압축돼 있고 왼쪽은 확장돼 있다.
- 그래도 LLE는 매니폴드를 모델링하는 데 잘 동작한다.

# 지역 선형 임베딩(LLE)

1. 알고리즘이 각 훈련 샘플  $x(i)$ 에 대해 가장 가까운  $k$ 개의 샘플 찾기
  2. 이웃에 대한 선형 함수로  $x(i)$ 를 재구성
    - $x(i)$ 와  $\sum_{j=1}^m w_{i,j} x(j)$  사이의 제곱 거리가 최소가 되는  $w_{i,j}$  찾기
    - $x(j)$ 가  $x(i)$ 의 가장 가까운  $k$ 개 이웃 중 하나가 아닐 경우에는  $w_{i,j} = 0$ 이 됨.
- $W$ : 가중치  $w_{i,j}$ 를 모두 담은 가중치 행렬
  - $\hat{W} = \operatorname{argmin} \sum_{i=1}^m (x(i) - \sum_{j=1}^m w_{i,j} x(j))^2$
  - [조건] 
$$\begin{cases} w_{i,j} = 0 & (x(j) \text{가 } x(i) \text{의 최근접 이웃 } k \text{개 중 하나가 아닐 때}) \\ \sum_{j=1}^m w_{i,j} = 1 & (i = 1, 2, \dots, m \text{ 일 때}) \end{cases}$$
  - $\hat{W}$ 은 훈련 샘플 사이에 있는 지역 선형 관계를 담고 있다.

# 지역 선형 임베딩(LLE)

- 가능한 이 관계가 보존되도록 훈련 샘플을 d차원 공간( $d < n$ )으로 매핑
- 만약  $z(i)$ 가 d차원 공간에서  $x(i)$ 의 상(image)이라면 가능한 한  $z(i)$ 와  $\sum_{j=1}^m \widehat{w}_{i,j} z(j)$  사이의 거리가 최소화되어야 함
  - > 제약이 없는 최적화 문제로 바꾸어 줌
- 첫 번째 단계와는 반대로 가중치를 고정하고 저차원의 공간에서 샘플 이미지의 최적 위치를 찾는다.
- $Z$  = 모든  $z(i)$ 를 포함하는 행렬
- $Z = \operatorname{argmin} \sum_{i=1}^m (z(i) - \sum_{j=1}^m w_{i,j} z(j))^2$

# 지역 선형 임베딩(LLE)

- LLE 구현의 계산 복잡도
  - $k$ 개의 가장 가까운 이웃 찾기  
:  $O(m \log(m) n \log(k))$
  - 가중치 최적화  
:  $O(mnk^3)$
  - 저차원 표현 만들기  
:  $O(dm^2)$   
:  $m^2$  때문에 이 알고리즘을 대량의 데이터셋에 적용하기는 어렵다.

## 8-6. 다른 차원 축소 기법



# 랜덤 투영

- 선형 투영을 사용해 데이터를 저차원 공간으로 투영
- 실제로 거리를 잘 보존한다.
- 차원 축소 품질은 샘플 수와 목표 차원수에 따라 다르다.
- 초기 차원수에는 의존적이지 않다.

# 다차원 스케일링(MDS)

- 샘플 간의 거리를 보존하면서 차원을 축소한다.
- 거리 기반 MDS
- 비거리 기반 MDS

# Isomap

- 각 샘플을 가장 가까운 이웃과 연결하는 식으로 그래프를 만든다.
- 샘플 간의 지오데식 거리를 유지하면서 차원을 축소한다.

# t-SNE

- 비슷한 샘플은 가까이, 비슷하지 않은 샘플은 멀리 떨어지도록 하면서 차원 축소
- 주로 시각화에 많이 사용됨, 특히 고차원 공간의 샘플 군집을 시각화할 때 사용됨
  - ex) MNIST 데이터셋을 2D로 시각화할 때

# 선형 판별 분석(LDA)

- 분류 알고리즘
- 훈련 과정에서 클래스 사이를 가장 잘 구분하는 축을 학습한다.
- 이 축은 데이터가 투영되는 초평면을 정의하는 데 사용할 수 있다.
- 투영을 통해 가능한 한 클래스를 멀리 떨어지게 유지시키므로 SVM 분류기 같은 다른 분류 알고리즘을 적용하기 전에 차원을 축소시키는 데 좋다.

# 다른 차원 축소 기법

