

11.2.2. 비지도 사전훈련

- 레이블 된 훈련데이터가 많지 않으면서 비슷한 작업에 대해 사전훈련 된 모델이 없을 때
- 레이블 된 훈련샘플을 모으는 비용 >> 레이블이 없는 훈련샘플을 모으는 비용
- 순서

1. 레이블이 없는 훈련데이터를 사용해 오토인코더, 생성적 적대 신경망(GAN; 17장) 같은 비지도학습 모델을 훈련

➔ 훈련데이터의 특성을 학습

2. 오토인코더, 생성적 적대 신경망의 하위층을 재사용(전이)

➔ 훈련데이터의 특성이 학습된 하위층

3. 새로운 작업에 맞는 출력층을 추가

4. 레이블 된 훈련샘플을 사용해 세밀하게 튜닝

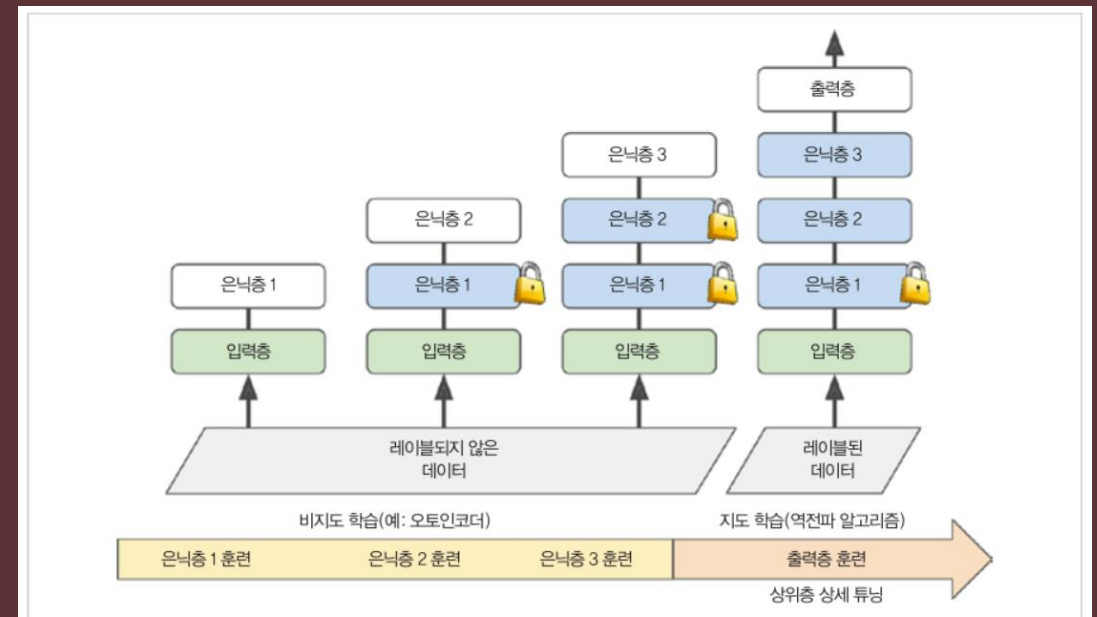


그림 11-5 비지도 훈련에서는 비지도 학습 기법으로 레이블이 없는 데이터(또는 전체 데이터)로 모델을 훈련합니다. 그 다음 지도 학습 기법을 사용하여 레이블된 데이터에서 최종 학습을 위해 세밀하게 튜닝합니다. 비지도 학습 부분은 그림처럼 한 번에 하나의 층씩 훈련하거나 바로 전체 모델을 훈련할 수도 있습니다.

11.2.3. 보조작업에서 사전훈련

- 레이블 된 훈련데이터가 많지 않은 경우
- 예시 1
 - 얼굴을 인식하는 신경망을 구축하고 싶다.
 - 그런데 개인별 얼굴 이미지가 별로 없다 (각 사람의 사진을 수백 장씩 모으기는 어려움)
 - 그러나 무작위로 많은 인물의 얼굴 이미지를 수집하기는 비교적 쉬움 (인터넷 구글)
 - 무작위 얼굴 이미지를 사용해 두 개의 얼굴 이미지가 같은 사람의 것인지 감지하는 신경망을 훈련
 - ➔ 이러한 신경망은 얼굴의 특성을 잘 감지하도록 학습이 되어있을 것임
 - 이 신경망의 하위층을 재사용(전이)해 얼굴을 분류하는 신경망을 훈련
- 예시 2: 자연어 처리(Natural Language Processing; NLP)
 - 수백만 개의 텍스트 문서로부터 레이블 된 데이터를 자동으로 생성 가능
 - 일부 단어를 랜덤하게 지우고 누락된 단어를 예측하는 모델을 훈련할 수 있음
 - What are you saying? 이라는 문장 → What ____ you saying? 과 같이 단어 지우기 → 누락된 단어를 예측하는 신경망 훈련 → 이 신경망의 하위층을 다른 비슷한 작업에 재사용(전이)

참고. 자기지도 학습 (Self-supervised learning)

- 데이터에서 스스로 레이블을 생성하고, 지도 학습 기법으로 레이블 된 데이터를 훈련하는 방법

고속 옵티마이저



개요

- 큰 심층 신경망(DNN)의 훈련 속도는 느릴 수 있음 (수많은 파라미터들 때문)
- 훈련 속도를 높이는 방법
 1. 가중치 초기화
 2. 활성화 함수
 3. 배치 정규화
 4. 사전훈련 층 재사용
 5. 표준적인 경사하강법 옵티마이저 → 모멘텀 최적화, 네스테로프 가속경사, AdaGrad, RMSProp, Adam, Nadam

11.3.1. 모멘텀 최적화

- 표준적인 경사하강법: 경사면을 따라 일정한 크기로 조금씩 내려감

➔ 최적 해에 도달하는 데 시간이 오래 걸림

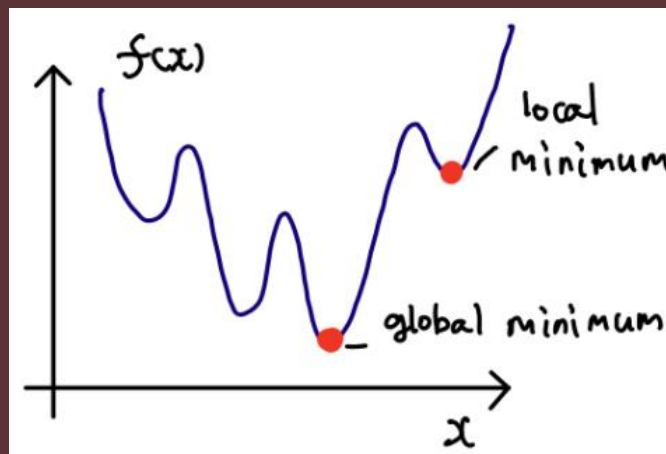
- 모멘텀 최적화: 이전 그래디언트(비용함수의 그래디언트)를 다음 학습에 반영
- 모멘텀 벡터 \mathbf{m} , 모멘텀 β , 학습률 η , 비용함수 J , 가중치 θ

- 모멘텀 β 가 0 ➔ 표준적인 경사하강법과 동일
- 모멘텀 β 가 1 ➔ 이전 그래디언트의 영향이 증가
- 일반적으로 모멘텀 β 는 0.9 사용
- 효과: Local minimum에 빠지지 않도록 도움

식 11-4 모멘텀 알고리즘²⁵

$$1. \mathbf{m} \leftarrow \beta \mathbf{m} - \eta \nabla_{\theta} J(\theta)$$

$$2. \theta \leftarrow \theta + \mathbf{m}$$



11.3.1. 모멘텀 최적화

- 구현

```
optimizer = keras.optimizers.SGD(lr = 0.001, momentum = 0.9)
```

