

인 공 지 능

[심층학습 기초 I]

본 자료는 해당 수업의 교육 목적으로만 활용될 수 있음.
일부 내용은 다른 교재와 논문으로부터 인용되었으며, 모든 저작권은 원 교재와 논문에 있음.

미리보기

■ 심층학습deep learning

- 다층 퍼셉트론multilayer perceptron에 은닉층hidden layer을 여러 개 추가하면 깊은 신경망deep neural networks이 됨
- 심층학습은 깊은 신경망의 학습
- 심층학습은 새로운 응용을 창출하고 인공지능 제품의 성능을 획기적으로 향상
→ 현대 기계학습을 주도

4.1 심층학습의 등장

- 4.1.1 심층학습의 성공 배경
- 4.1.2 특징 학습의 부각

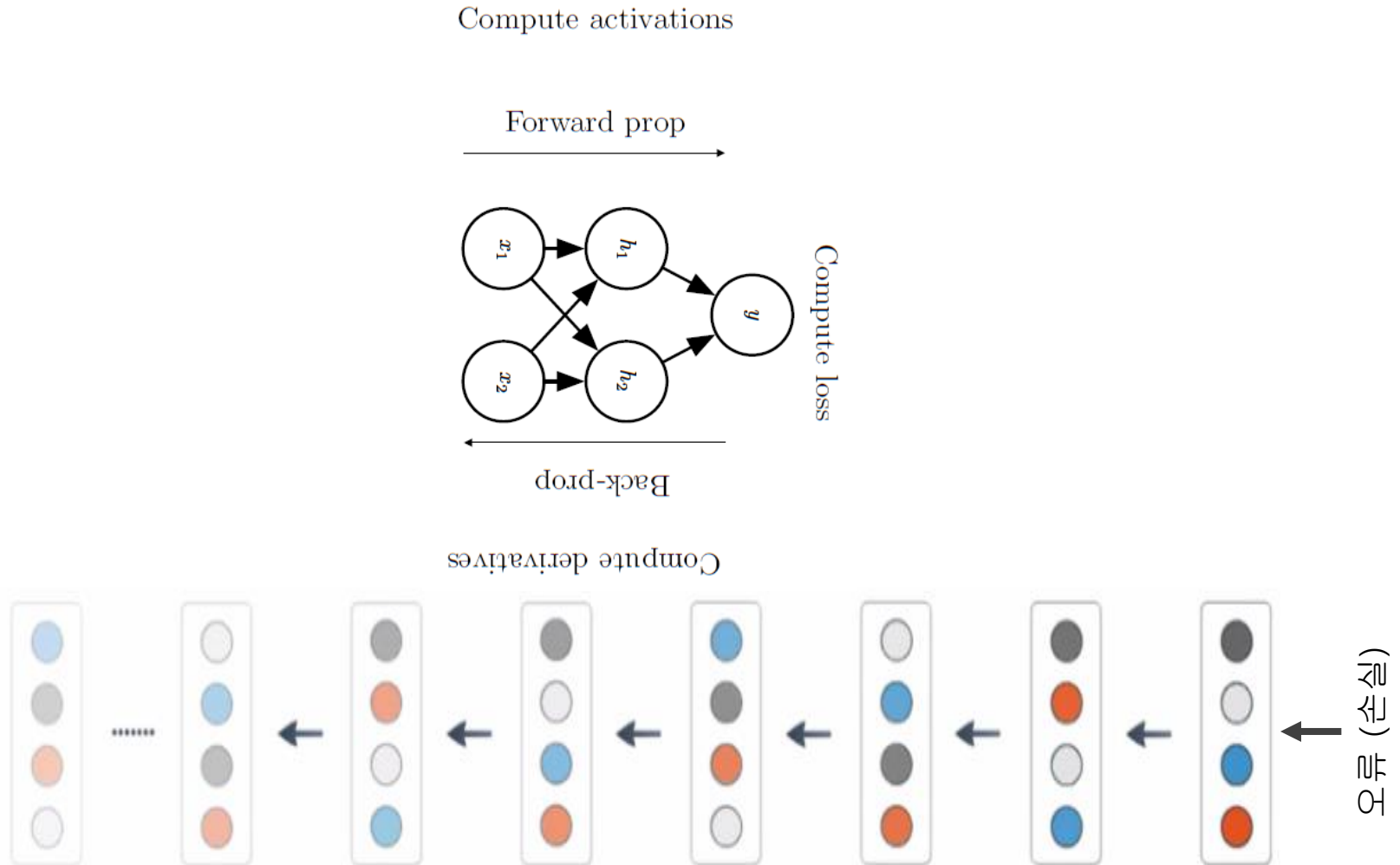
4.1 심층학습의 등장

■ 배경

- 1980년대에 이미 깊은 신경망 아이디어 등장
- 하지만 실현 불가능 (당시, 깊은 신경망은 학습이 안 됨)
 - 경사 소멸 vanishing gradient 문제
 - 작은 훈련집합
 - 과다한 연산과 시간 소요 (낮은 연산의 범용 컴퓨터, 값비싼 슈퍼컴퓨터)
- 일부 연구자들은 실망스러운 상황에서도 지속적인 연구
 - 학습률에 따른 성능 변화 양상
 - 모멘텀과 같은 최적 탐색 방법 모색
 - 은닉 노드 수에 따른 성능 변화
 - 데이터 전처리의 영향
 - 활성화함수의 영향
 - 규제 기법의 영향 등

4.1 심층학습의 등장

■ 경사 소멸 문제|gradient vanishing problem



4.1.1 심층학습의 성공 배경

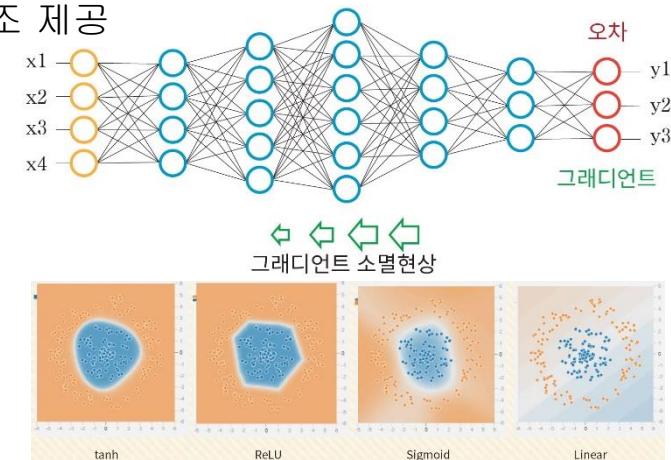
■ 이유

■ 혁신적 알고리즘 등장

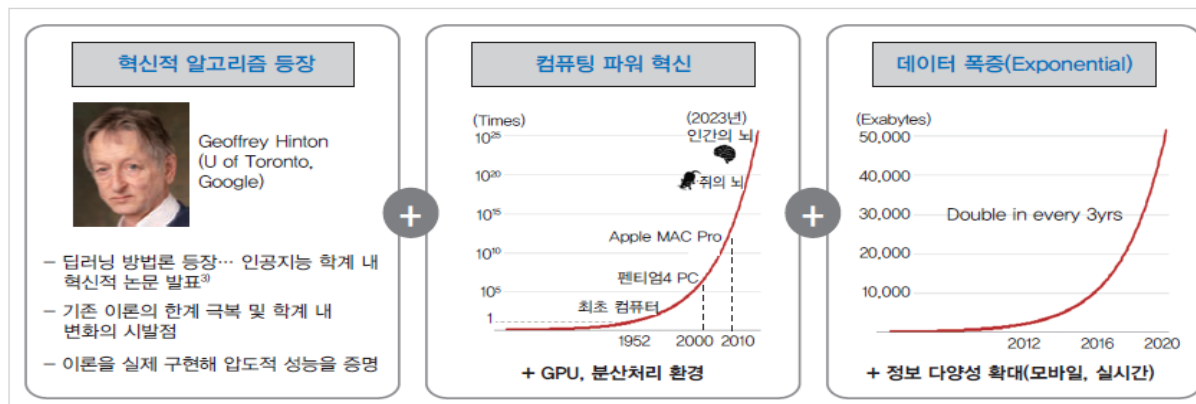
- 합성곱 신경망convolutional neural networks (CNN) 구조
 - 부분 연결과 가중치 공유를 통해서 효율적인 신경망 학습 구조 제공
- 경사 소멸 문제 해결을 위한 ReLU 활성화함수
- 과잉적합을 방지하는데 효과적인 다양한 규제 기법
- 층별 예비학습pretraining 기법 개발

■ 값싼 GPGPU의 등장

■ 학습 데이터 양과 질의 향상



[LG경제연구원]



4.1.2 표현 학습의 부각

■ 기계학습의 새로운 전환

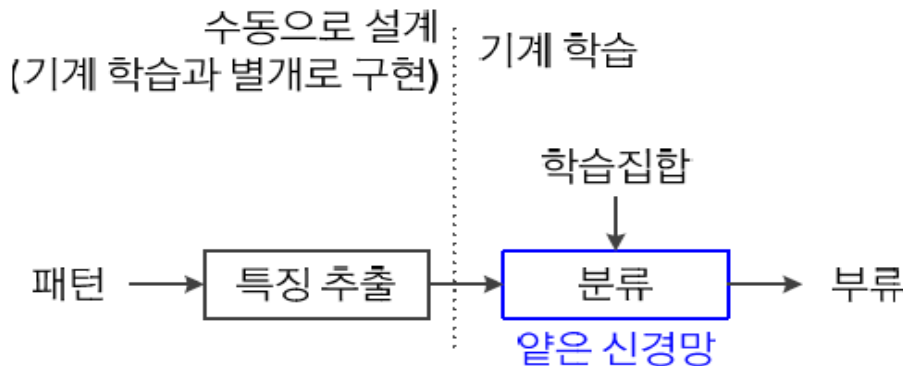
- 전통적인 다층 퍼셉트론

- 은닉층은 특징 추출기

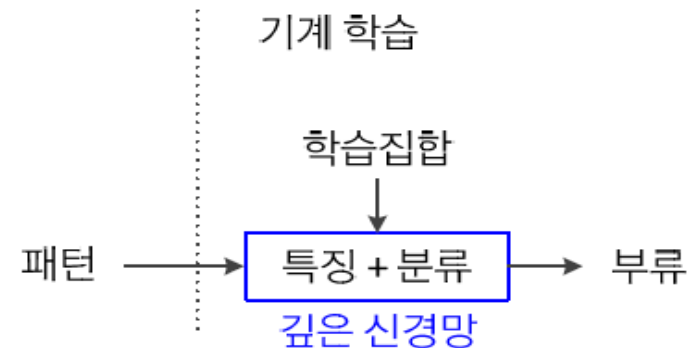
- 얇은 구조 (제한적 특징 추출)이므로 가공하지 않은 획득한 원래 패턴을 그대로 입력하면 낮은 성능
→ 따라서 사람이 수작업 특징을 선택(feature selection)하거나 추출(feature extraction)하여 신경망에 입력함

- 현대 기계학습 (심층학습)

- 학습에 의해 자동적으로 데이터로부터 특징(data-driven features) 추출 ← 표현 학습(representation learning)
– 특징 벡터를 신경망의 입력 ← 종단간 학습(end-to-end learning)



(a) 고전적 패러다임에서의 수작업 특징

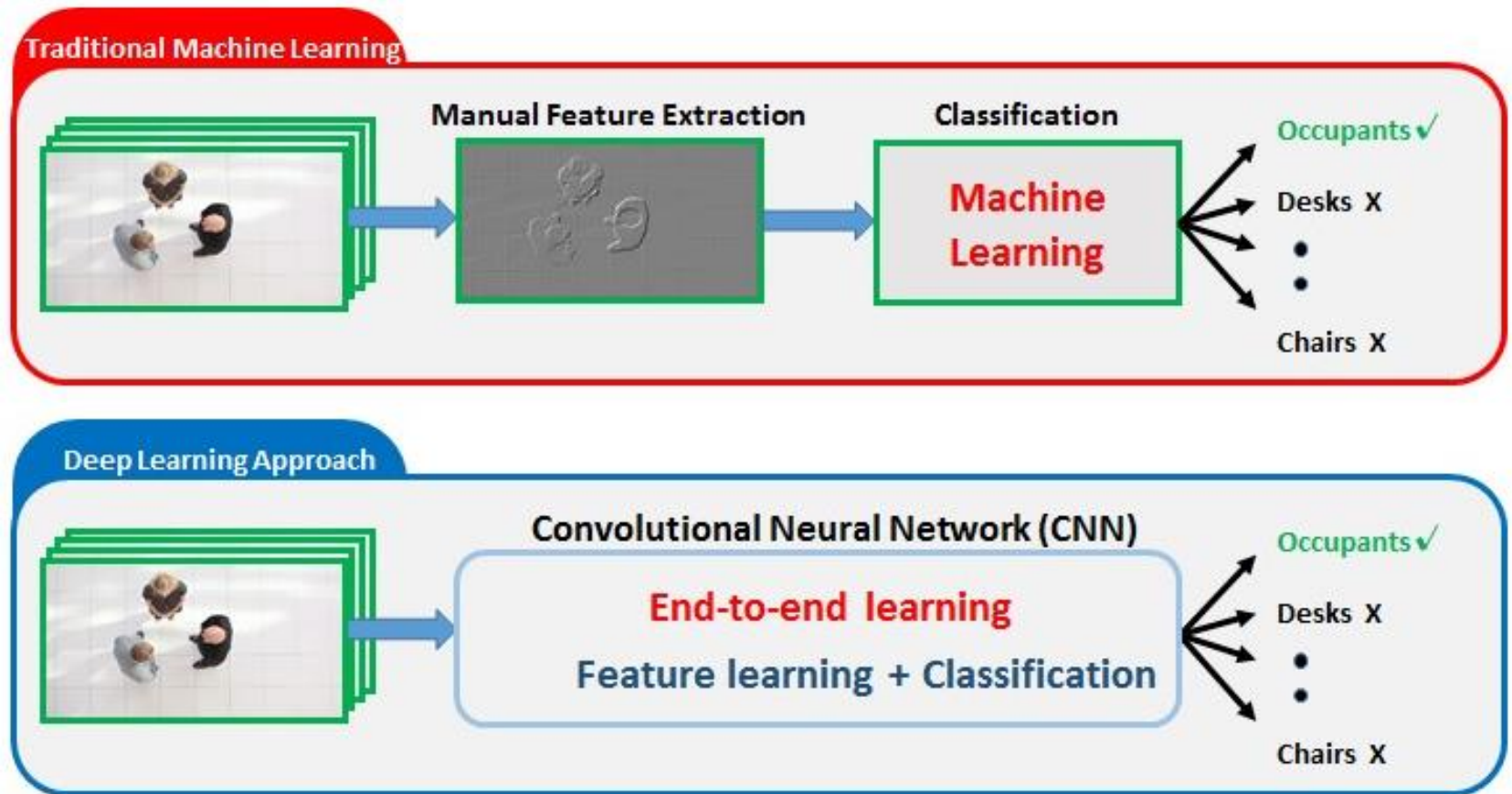


(b) 딥러닝에서의 특징 학습

그림 4-1 기계 학습의 패러다임 변화

4.1.2 표현 학습의 부각

■ 고전적인 기계학습과 심층학습 비교 예



→ 깊은 신경망을 통한 계층적 표현 학습

4.1.2 표현 학습의 부각

■ 깊은 신경망의 표현 학습representation learning (또는 특징 학습feature learning)

- 낮은 단계 은닉층은 선이나 모서리와 같은 간단한 (저급) 특징 추출
- 높은 단계 은닉층은 추상적인 형태abstractive representation의 복잡한 (고급) 특징을 추출

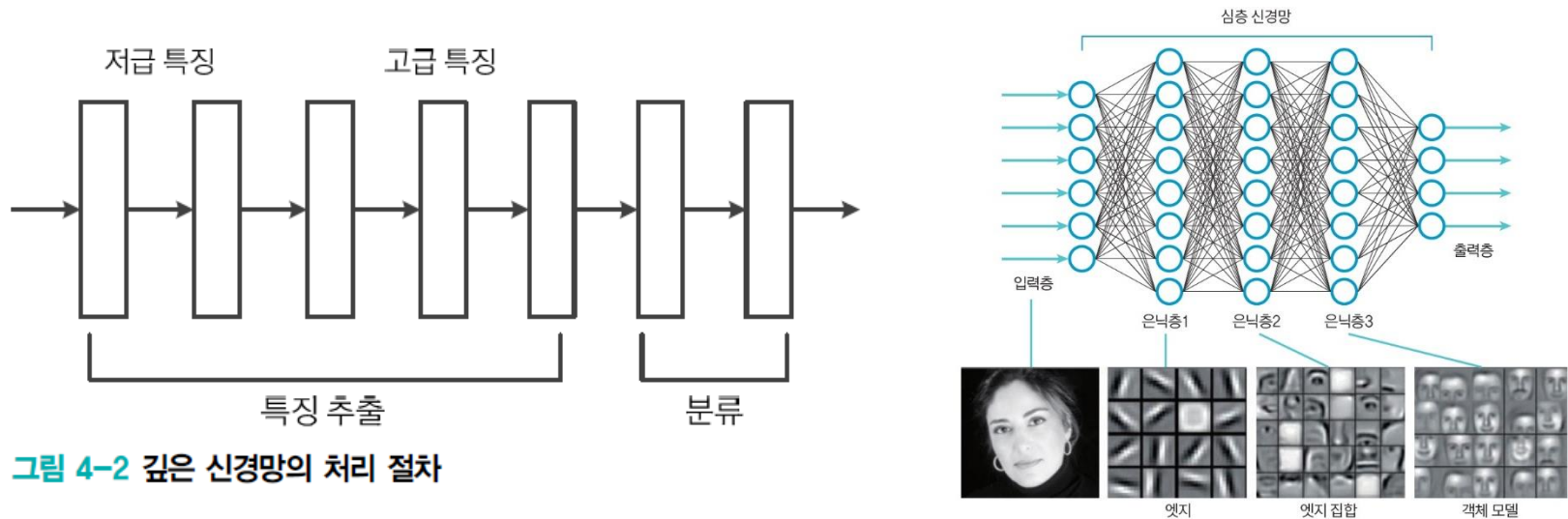


그림 4-2 깊은 신경망의 처리 절차

→ 표현 학습이 강력해짐에 따라 기존 응용에서 획기적인 성능 향상

- 영상 인식, 음성 인식, 언어 번역 등
- 새로운 응용 창출
 - 분류나 회귀 뿐 아니라 생성 모델이나 화소pixel 수준의 영상 분할segmentation
 - CNN과 LSTM의 혼합 학습 모델 (예, 자연 영상에 주석 달기 응용) 등이 가능해짐

4.2 깊은 다층 퍼셉트론 (깊은 신경망)

- 4.2.1 구조와 동작
- 4.2.2 학습

4.2.1 구조와 동작

■ 깊은 다층 퍼셉트론(DMLP 혹은 deep MLP)의 구조

- **입력** ($d+1$ 차원의 특징벡터)과 **출력** (c 개 분류)
- $L-1$ 개의 **은닉층** (입력층은 0번째 은닉층, 출력층은 L 번째 은닉층으로 간주)
 - l 번째 은닉층의 노드 수를 n_l 로 표기

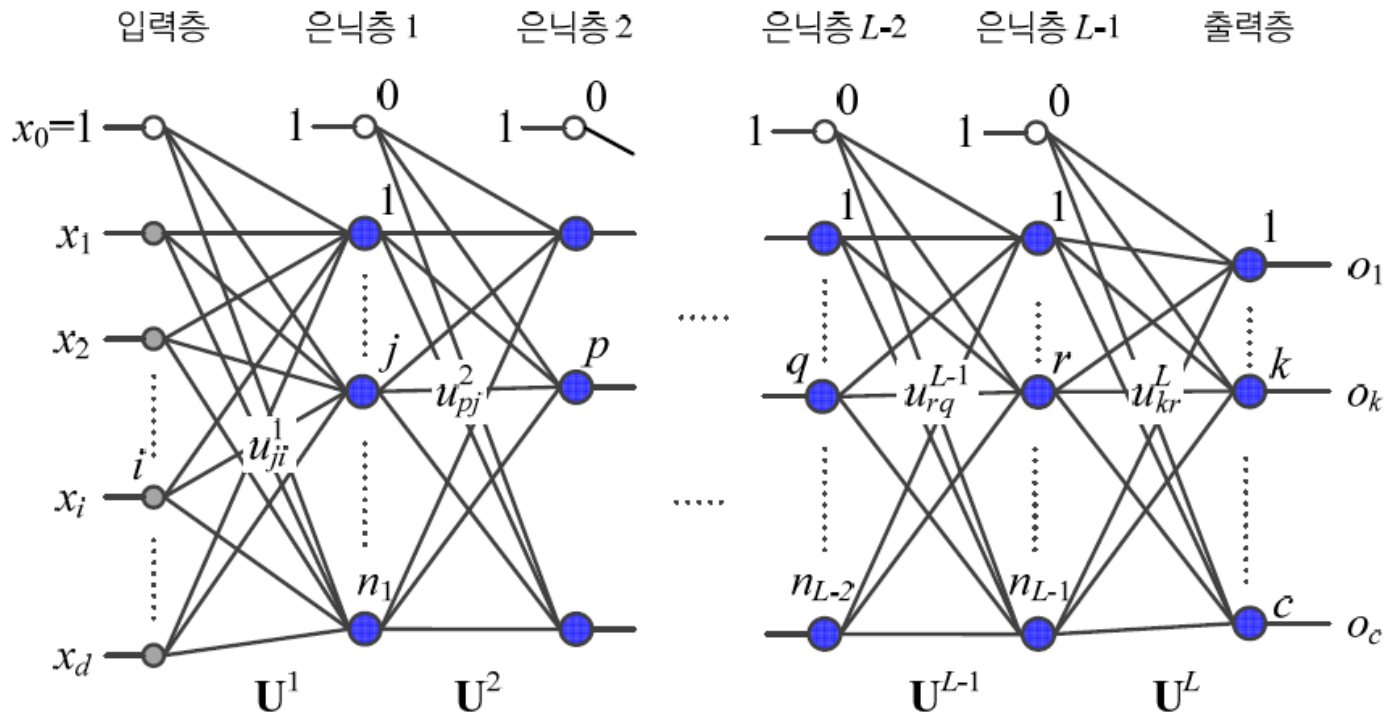


그림 4-3 깊은 MLP(DMLP)의 구조

4.2.1 구조와 동작

■ DMLP의 가중치 행렬

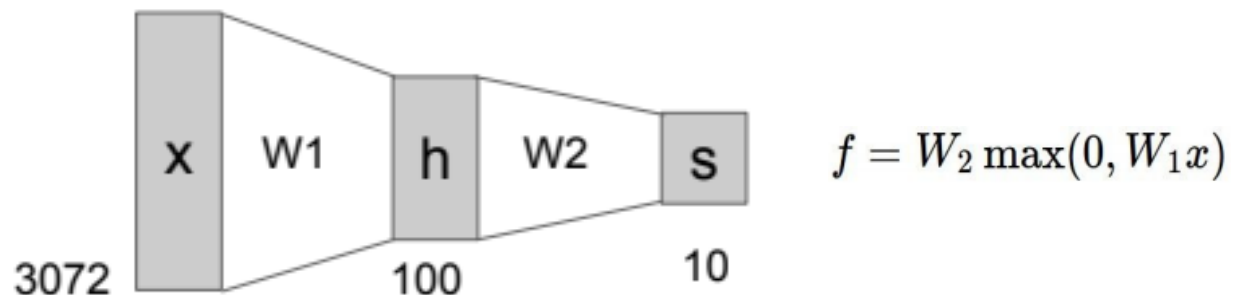
- u_{ji}^l 은 $l-1$ 번째 층의 i 번째 노드와 l 번째 층의 j 번째 노드를 연결하는 가중치
- $l-1$ 번째 층과 l 번째 층을 연결하는 가중치는 총 $(n_{l-1} + 1)n_l$ 개

$$\text{가중치 행렬: } \mathbf{U}^l = \begin{pmatrix} u_{10}^l & u_{11}^l & \cdots & u_{1n_{l-1}}^l \\ u_{20}^l & u_{21}^l & \cdots & u_{2n_{l-1}}^l \\ \vdots & \vdots & \ddots & \vdots \\ u_{n_l0}^l & u_{n_l1}^l & \cdots & u_{n_ln_{l-1}}^l \end{pmatrix}, l = 1, 2, \dots, L \quad (4.1)$$

■ DMLP의 동작

- **MLP의 동작**을 나타내는 식을 보다 많은 단계로 **확장**한 것

$$\mathbf{o} = \mathbf{f}(\mathbf{x}) = \mathbf{f}_L \left(\cdots \mathbf{f}_2(\mathbf{f}_1(\mathbf{x})) \right) \quad (4.2)$$



4.2.1 구조와 동작

■ 동작(전방 계산 forward pass)을 구체적으로 쓰면,

- 입력층의 특징 벡터를 내부 표현으로 바꾸어 쓰면,

$$\mathbf{z}^0 = (z_0, z_1, z_2, \dots, z_{n_0})^T = (1, x_1, x_2, \dots, x_d)^T \quad (4.3)$$

- l 번째 층의 j 번째 노드가 수행하는 연산

l 번째 은닉층의 j 번째 노드의 연산:

$$z_j^l = \tau_l(s_j^l)$$

이때 $s_j^l = \mathbf{u}_j^l \mathbf{z}^{l-1}$ 이고,

$$\mathbf{z}^{l-1} = (1, z_1^{l-1}, z_2^{l-1}, \dots, z_{n_{l-1}}^{l-1})^T, \quad \mathbf{u}_j^l = (u_{j0}^l, u_{j1}^l, \dots, u_{jn_{l-1}}^l)$$

(4.4)

- 행렬 표기를 이용하여 l 번째 층의 연산 전체를 쓰면,

$$l\text{번째 층의 연산: } \mathbf{z}^l = \boldsymbol{\tau}_l(\mathbf{U}^l \mathbf{z}^{l-1}), \quad 1 \leq l \leq L \quad (4.5)$$

4.2.2 학습

■ DMLP 학습은 기존 MLP 학습과 유사

- DMLP는 경사도 계산과 가중치 갱신을 더 많은 단계(층)에 걸쳐 수행

■ 오류 역전파 알고리즘

- L 번째 층(출력층)의 경사도 계산

$$\delta_k^L = \tau'_L(s_k^L)(y_k - o_k), \quad 1 \leq k \leq c \quad (4.6)$$

$$\frac{\partial J}{\partial u_{kr}^L} = -\delta_k^L z_r^{L-1}, \quad 0 \leq r \leq n_{L-1}, 1 \leq k \leq c \quad (4.7)$$

- $l+1$ 번째 층의 정보를 이용하여 l 번째 층의 경사도 계산($l=L-1, L-2, \dots, 1$)

$$\delta_j^l = \tau'_l(s_j^l) \sum_{p=1}^{n_{l+1}} \delta_p^{l+1} u_{pj}^{l+1}, \quad 1 \leq j \leq n_l \quad (4.8)$$

$$\frac{\partial J}{\partial u_{ji}^l} = -\delta_j^l z_i^{l-1}, \quad 0 \leq i \leq n_{l-1}, 1 \leq j \leq n_l \quad (4.9)$$

4.2.2 학습

- 식 (4.6)~식 (4.9)의 오류 역전파 공식을 이용한 학습 알고리즘

알고리즘 4-1 DMLP를 위한 미니배치 스토캐스틱 경사 하강법

입력: 훈련집합 \mathbb{X} 와 \mathbb{Y} , 학습률 ρ , 미니배치 크기 t

출력: 가중치 행렬 $\mathbf{U}^l, l = 1, 2, \dots, L$

```
1   $\mathbf{U}^l, l = 1, 2, \dots, L$ 을 초기화한다.
2  repeat
3       $\mathbb{X}$ 와  $\mathbb{Y}$ 에서  $t$ 개의 샘플을 무작위로 뽑아 미니배치  $\mathbb{X}'$ 와  $\mathbb{Y}'$ 를 만든다.
4      for ( $l=1$  to  $L$ )  $\Delta \mathbf{U}^l = \mathbf{0}$ 
5          for ( $\mathbb{X}'$ 의 샘플 각각에 대해)
6              현재 처리하는 샘플을  $\mathbf{x} = (x_0, x_1, x_2, \dots, x_d)^T$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_c)^T$ 라 표기한다.
7               $x_0, z_0^1, z_0^2, \dots, z_0^L$ 을 1로 설정한다.
8              // 전방 계산
9               $\mathbf{x}$ 를  $\mathbf{z}^0$ 에 대입한다. // 식 (4.3)
10             for ( $l=1$  to  $L$ ) // 왼쪽 층에서 오른쪽 층으로 진행하면서 전방 계산
11                 for ( $j=1$  to  $n_l$ ) // 각 노드에 대해
12                      $s_j^l = \mathbf{u}_j^l \mathbf{z}^{l-1}$  // 식 (4.4)
13                      $z_j^l = \tau_l(s_j^l)$  // 식 (4.4)
14                 // 오류 역전파의 단계 1: 그레이디언트 계산
15                 for ( $k=1$  to  $c$ )  $\delta_k^L = \tau'_L(s_k^L)(y_k - o_k)$  // 식 (4.6)
16                 for ( $k=1$  to  $c$ ) for ( $r=0$  to  $n_{L-1}$ )  $\Delta u_{kr}^L = \Delta u_{kr}^L + (-\delta_k^L z_r^{L-1})$ 
17                 for ( $l=L-1$  to 1) // 오른쪽 층에서 왼쪽 층으로 진행하면서 오류 역전파
18                     for ( $j=1$  to  $n_l$ )  $\delta_j^l = \tau'_l(s_j^l) \sum_{p=1}^{n_{l+1}} \delta_p^{l+1} u_{pj}^{l+1}$  // 식 (4.8)
19                     for ( $j=1$  to  $n_l$ ) for ( $i=0$  to  $n_{l-1}$ )  $\Delta u_{ji}^l = \Delta u_{ji}^l + (-\delta_j^l z_i^{l-1})$ 
20                 // 오류 역전파의 단계 2: 가중치 갱신
21             for ( $l=L$  to 1)
22                 for ( $j=1$  to  $n_l$ ) for ( $i=0$  to  $n_{l-1}$ )  $u_{ji}^l = u_{ji}^l - \rho \left( \frac{1}{t} \right) \Delta u_{ji}^l$ 
23 until (멈춤 조건)
```

4.2.2 학습

■ 역사적 고찰

- 주요 알고리즘의 개선

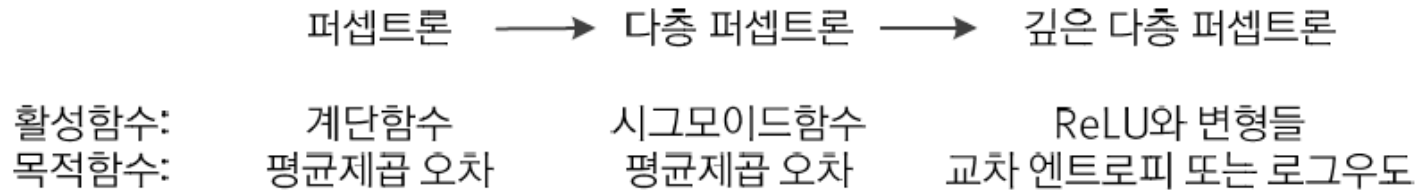


그림 4-4 다층 퍼셉트론의 역사적 발전 양상

- 합성곱 신경망(convolutional neural networks(CNNs)의 부상
 - 예) MNIST 인식 경쟁
 - 2010년 784-2500-2000-1500-1000-500-10 구조의 DMLP: 0.35% 오류율
 - 2011년 CNN: 0.35% 오류율
 - 2012년 35개 CNN을 이용한 앙상블 모델: 0.23% 오류율
 - 예) ILSVRC 사진 인식 경쟁: CNN이 DMLP보다 확연히 우월

4.6 심층학습은 왜 강력한가?

■ 종단간^{end-to-end} 최적화된 학습 가능

- 고전적인 방법에서는
 - [그림 4-37]처럼 분할, 특징 추출, 분류를 따로 구현한 다음 이어 붙임
 - 사람의 직관에 따르므로 성능 한계
 - 인식 대상이 달라지면 새로 설계해야 함

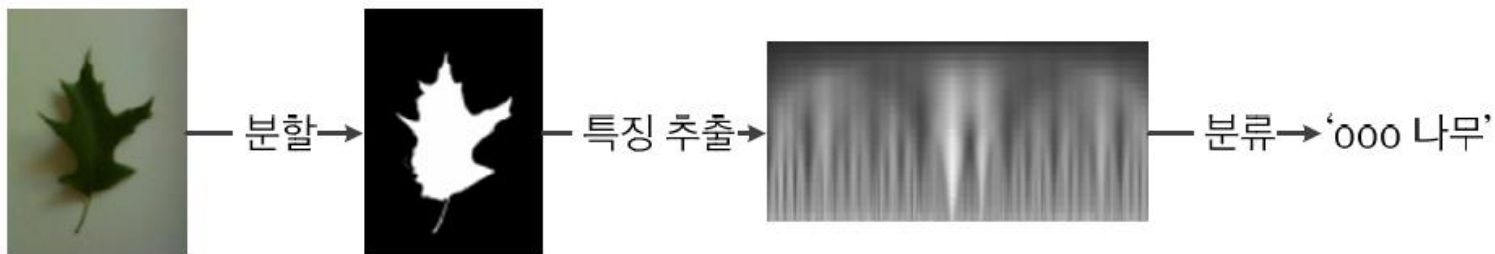


그림 4-37 여러 단계를 따로 설계 구현하는 고전적인 접근방식(나뭇잎 인식 사례)

- 심층학습은 전체 깊은 신경망을 동시에 최적화 (종단간 학습이라 부름)

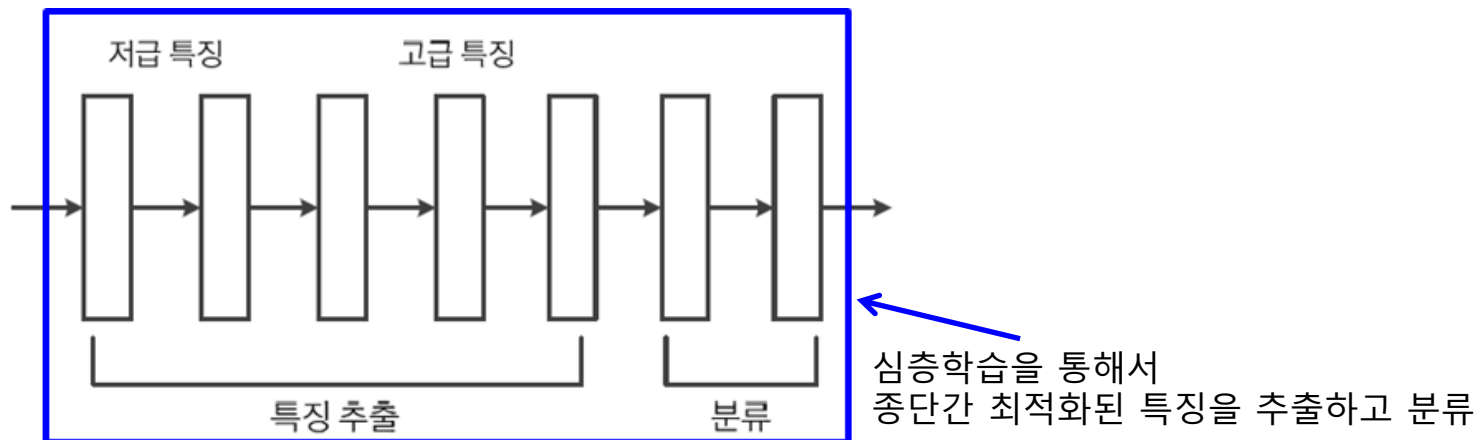


그림 4-2 깊은 신경망의 처리 절차

4.6 심층학습은 왜 강력한가?

■ 깊이^{depth}의 중요성

- 점선은 20개 노드를 가진 은닉층 하나 짜리 신경망 → 더 정교한 분할 → 과대적합 문제
- 실선은 각각 10개 노드를 가진 은닉층 두 개 짜리 신경망

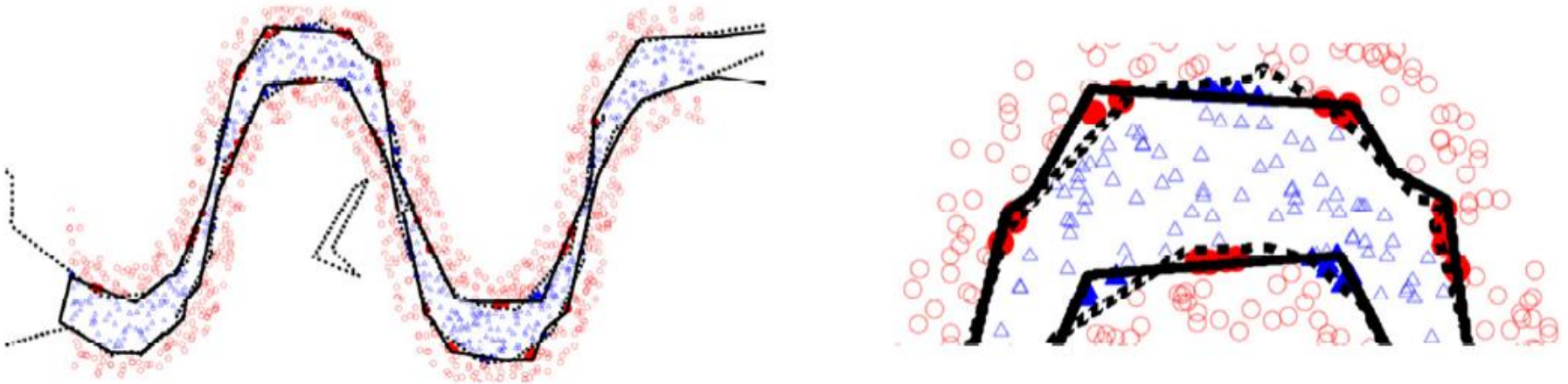


그림 4-38 은닉층의 개수가 늘어남에 따른 표현력 증가

4.6 심층학습은 왜 강력한가?

■ 계층적 특징 hierarchical features

- [그림 4-40]은 ImageNet으로 학습한 특징 맵 feature map → 추출된 특징 표현
 - 깊은 신경망에서는 층의 역할이 잘 구분됨
 - 반면 얇은 신경망은 하나 또는 두 개의 은닉층이 여러 형태의 특징을 모두 담당

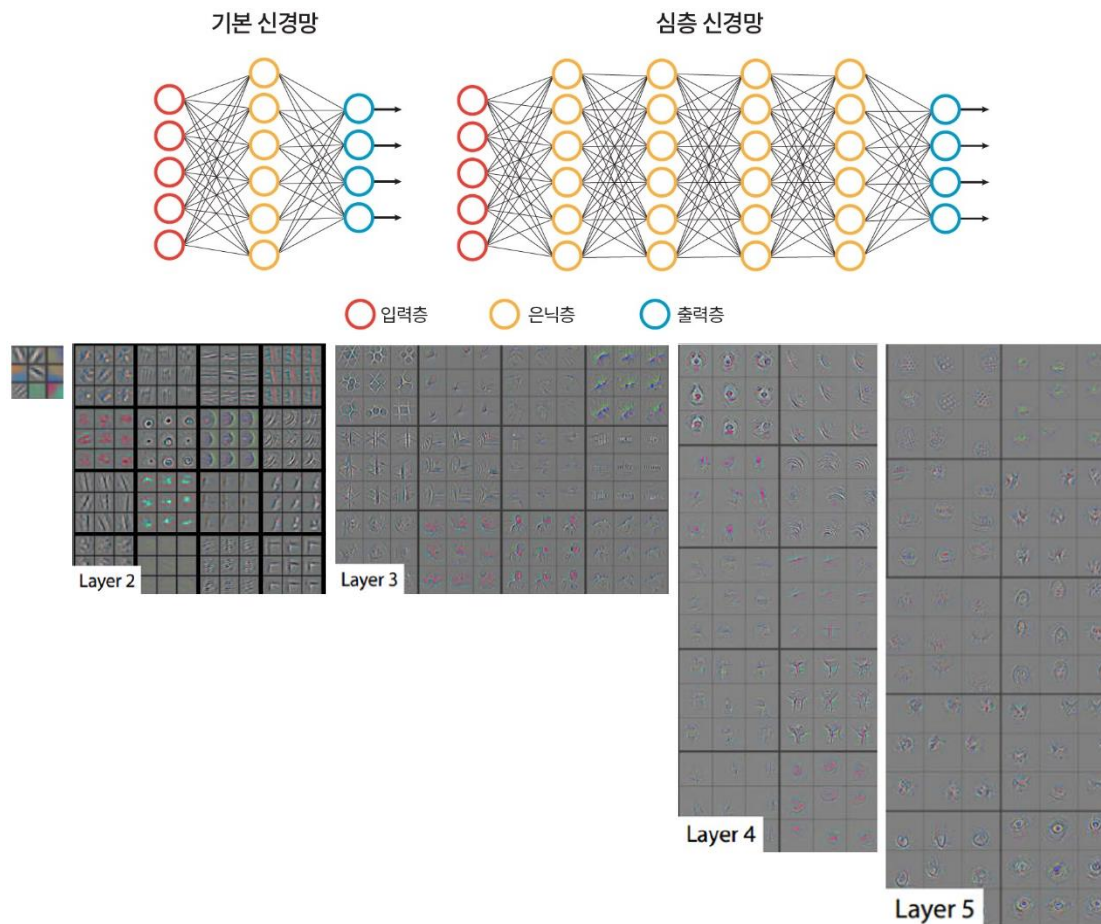
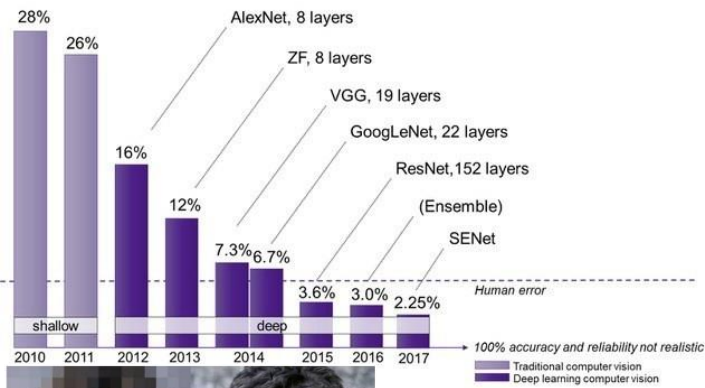


그림 4-40 CNN의 계층적 특징 추출

4.6 심층학습은 왜 강력한가?

■ 최신 심층학습 성능



KorQuAD 1.0

Question Answering Dataset

SQuAD 2.0

The Stanford Question Answering Dataset

Leaderboard

SQuAD 2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Ji et al. '18)	86.831	89.452
1	SA-Net on Albert (ensemble) QIANXIN	90.724	93.011
2	SA-Net-V2 (ensemble) QIANXIN	90.679	92.948
2	Retro-Reader (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09494v2	90.578	92.978
3	ELECTRA+ALBERT+EntitySpanFocus (ensemble) SRCB_DMIL	90.442	92.839

Leaderboard

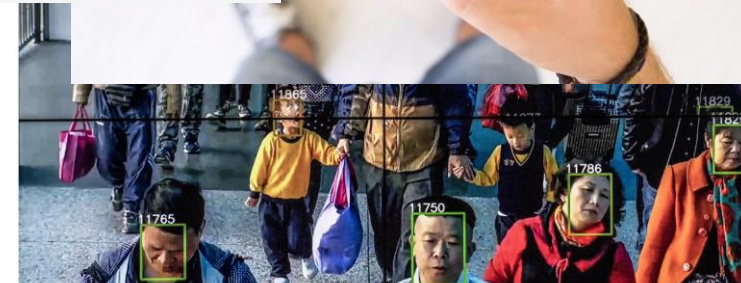
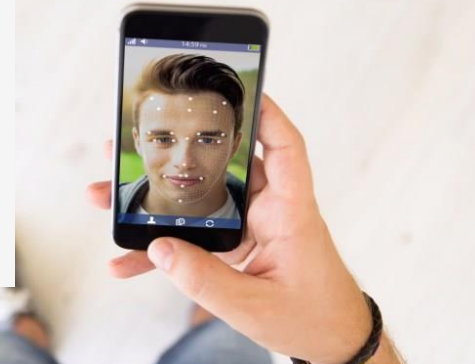
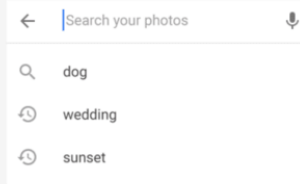
KorQuAD 1.0의 Test set으로 평가한 Exact Match(EM) 및 F1 score입니다.

Rank	Reg. Date	Model	EM	F1
-	2018.10.17	Human Performance	80.17	91.20
1	2020.01.08	SkERT-Large (single model) Skelter Labs	87.66	95.15



Find your photos faster

Your photos are organized and searchable by the places and things in them – no tagging required. Just search "dog" to find all the photos of your pup.



Collage of some medical imaging applications in which deep learning has achieved state-of-the-art results.

From top-left to bottom-right:

1. mammographic mass classification
2. segmentation of lesions in the brain,
3. leak detection in airway tree segmentation,
4. diabetic retinopathy classification
5. prostate segmentation,
6. nodule classification,
7. breast cancer metastases detection,
8. skin lesion classification
9. bone suppression