

연습문제

- 1 [그림 1-7]의 선형분리 불가능한 상황을 선형분리 가능한 상황으로 변환해 주는 자신의 함수를 고안하여 제시하시오.
- 2 다음 훈련집합을 3차원 공간에 그리고, 선형분리 가능 여부와 그 이유를 제시하시오.
$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$
$$y_1 = 1, y_2 = 1, y_3 = 1, y_4 = -1, y_5 = -1, y_6 = -1$$
- 3 [예제 1-1]에서 $\theta = (w, b)^T = (-0.1, 4.0)^T$ 일 때 목적함수의 값을 계산하시오. b 를 4.0으로 고정한다면 w 를 얼마로 해야 목적함수가 최저가 되는지 기술하시오.
- 4 [그림 1-10]은 4차원 특징 공간을 여러 개의 2차원 조합으로 나누어 그린 것이다. 4개 특징 중에 2개만 사용해야 한다면 어느 쌍을 사용할지 개략적으로 설명하시오.
- 5 종이에 “대한민국”을 10개 쓰시오. 한 번은 최선을 다해 비슷한 모양이 되도록 쓰고, 다음에는 최대한 서로 다른 모양이 되도록 쓰시오. 결과를 가지고 필기 한글 인식 문제의 난이도에 대한 견해를 제시하시오.
- 6 [표 1-1]에서 새로운 기준을 개발하고 사람과 기계의 학습을 비교하시오.
- 7 <http://yann.lecun.com/exdb/mnist/>에 접속하면 [Ciresan2012]가 MNIST 필기 숫자 데이터베이스에서 현재 최고 성능을 보인 논문이라는 사실을 알 수 있다. 이 논문은 테스트 샘플 10,000자 중 틀린 23자를 제시한다. 논문을 참조하여 틀린 샘플을 제시하고, 이들 오류에 대한 자신의 견해를 제시하시오.

8 UCI 리퍼지토리에서 특징 벡터의 차원이 100 이하인 데이터베이스 2개, 100보다 크고 1000 이하인 것 2개, 1000보다 큰 것 2개를 선정하여 부류 개수, 특징 개수, 샘플 개수를 조사하시오.

9 뇌 질환과 관련한 데이터베이스를 조사하고, 많이 쓰이는 2가지를 골라 데이터베이스 내용에 대해 기술하시오.

10 이 책의 참고문헌을 주의 깊게 살펴보면 기계 학습을 선도하는 논문은 주로 저널보다 학술 대회에서 발표된다는 사실을 알 수 있다. 예를 들어, NIPS^{Neural Information Processing Systems}와 ICML^{International Conference on Machine Learning}은 기계 학습의 혁신적인 아이디어와 혁신적인 실험 결과가 주로 발표되는 주요 학술대회이다. 이러한 현상은 컴퓨터 공학의 학문 특성에 따라 일어나는데, 어떤 특성인지 조사하고 자신의 의견을 덧붙여 기술하시오.

Hint 위키피디아에서 "computer science" 참조

11 앱인벤터AppInventor^는 안드로이드 앱을 개발하는 데 쓰는 언어인데, 비전공자들이 쉽게 배울 수 있는 장점이 있다. 앱인벤터는 기계 학습과 관련된 컴포넌트를 몇 가지 제공하는데, 성능이 뛰어나다. 이들 컴포넌트를 조사하시오. 또한 성능 평가 기준을 세우고 간단한 테스트용 앱을 제작하여 컴포넌트의 성능을 평가 분석하시오.

Hint <http://appinventor.chonbuk.ac.kr> 참조

12 다음은 기계 학습으로 제작된 인공지능 시스템인데, 웹을 통해 데모 서비스를 제공한다. 성능 기준을 스스로 개발하고 성능을 측정 분석하시오. 특히 사람의 능력과 전주었을 때 우열에 대한 자신의 생각을 제시하시오.

- 영상 주석 달기(<http://deeplearning.cs.toronto.edu/i2t>)
- 자연 영상 인식(<http://demo.caffe.berkeleyvision.org/>)
- 필기 흉내내기(<http://www.cs.toronto.edu/~graves/handwriting.html>)
- 구글 번역(<https://translate.google.co.kr/>)
- 구글 음성 인식(<http://google.com>)
- 사진을 화풍으로 변환(<http://deepdreamgenerator.com>)

13 1.8.3절의 사회적 전망을 참고하여, 인공지능의 현재와 미래 전망에 대한 자신의 견해를 밝히시오.

연습문제

1 $A = \begin{pmatrix} 1 & -2 & 3 & 5 \\ 2 & 2 & -1 & 0 \\ 3 & 0 & 1 & 2 \\ 1 & 0 & 2 & 0 \end{pmatrix}$ 의 $2A, A^T, A^{-1}$ 을 쓰시오. 또한 A 의 계수를 쓰시오.

Hint Matlab 또는 Python과 같은 도구를 사용하시오.

2 $A = \begin{pmatrix} 1 & -2 \\ 2 & 2 \end{pmatrix}$ 의 $A^{\frac{1}{2}}$ 과 $A^{-\frac{1}{2}}$ 을 구하시오.

Hint $M = A^{\frac{1}{2}}$ 이라면 $M * M = A$ 이고, $M = A^{-\frac{1}{2}}$ 이라면 $M * M = A^{-1}$ 이다. Matlab 또는 Python과 같은 도구를 사용하시오.

3 [예제 2-4]의 [그림 2-11(a)]와 [그림 2-11(c)]에서 행렬식이 역할을 제대로 수행하는지 확인하시오.

4 변수가 4개인 다음 2차 다항식 2개를 [예제 2-1]처럼 행렬로 표시하시오. 행렬을 더함으로써 다항식 덧셈 $f_1(x) + f_2(x)$ 를 수행할 수 있다는 사실을 이 예제를 이용하여 보이시오.

$$f_1(x) = -2x_1x_1 - 1.5x_1x_2 + 3.1x_1x_4 + 2x_2x_1 + 2x_2x_2 - 6x_2x_4 - 2.1x_3x_1 + 3x_3x_3 + 2x_4x_1 + 3x_4x_2 - 2x_4x_4 + 1.2x_1 + 3.3x_2 - 4x_4 + 4.2$$

$$f_2(x) = 3x_1x_1 + 3.5x_1x_3 - 3.1x_1x_4 + 2.1x_2x_1 - 2x_2x_2 - 2x_2x_4 - 2.5x_3x_1 + 3x_3x_3 + 3x_4x_1 - 6x_4x_2 + 2x_4x_3 + 2.2x_1 + 2.3x_2 + 3.3x_3 - 2x_4 + 5.1$$

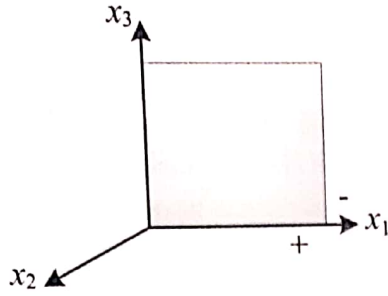
5 norms을 계산하시오.

(1) $x = (3 \ -4 \ -1.2 \ 0 \ 2.3)^T$ 의 1차, 2차, 3차 norms과 최대 norm

(2) $\begin{pmatrix} 2 & 1 \\ 1 & 5 \\ 4 & 1 \end{pmatrix}$ 의 프로베니우스 norm

6 [예제 2-3]에서 $T = 2.0$ 으로 바뀌었다고 가정하고 [그림 2-4(b)]를 새로 그리시오. T 가 변함에 따라 어떤 변화가 나타나는지 설명하시오.

- 7 아래 그림은 x_1 과 x_3 축이 이루는 평면이 결정평면인 상황이다. 이에 해당하는 퍼셉트론을 [그림 2-4(a)]처럼 제시하시오.



- 8 $m = 8$ 일 때 $p = 0.1, 0.5, 0.8$ 각각의 이항 분포를 그림으로 그리시오.

- 9 옷놀이에서 $P(\text{red}) = 0.4, 0.5, 0.6$ 인 3가지 경우에 대해 답하시오.

- (1) 각 경우의 확률분포를 구하시오.
- (2) 각 경우의 엔트로피를 구하시오.
- (3) $P(\text{red}) = 0.4$ 와 $P(\text{red}) = 0.5$ 의 교차 엔트로피와 $P(\text{red}) = 0.4$ 와 $P(\text{red}) = 0.6$ 의 교차 엔트로피를 구하시오. 어느 것이 큰지 확인하고 그 이유를 설명하시오.
- (4) $P(\text{red}) = 0.4$ 와 $P(\text{red}) = 0.5$ 의 KL 다이버전스와 $P(\text{red}) = 0.4$ 와 $P(\text{red}) = 0.6$ 의 KL 다이버전스를 구하시오. 어느 것이 큰지 확인하고 그 이유를 설명하시오.

- 10 특징 벡터가 3차원이면 퍼셉트론은 매개변수 4개를 가진다. [그림 2-4(a)]가 이 경우에 해당한다. 모든 매개변수가 $[-1, 1]$ 범위에 있고 이 범위를 0.0001 간격으로 나누어 탐색한다고 가정한다.

- (1) 탐색해야 하는 점의 개수를 쓰시오. 점 하나를 평가하는 데 0.01밀리초가 걸린다면, 학습에 걸리는 시간은?
- (2) [그림 1-6]에 있는 Wine 데이터베이스의 특징 벡터는 13차원이다. 앞의 가정을 그대로 적용한다면 Wine 데이터베이스를 학습할 때 점의 개수와 학습에 걸리는 시간을 쓰시오.

- 11 $f: \mathbb{R}^3 \mapsto \mathbb{R}^2$ 인 다음 벡터 함수의 야코비안 행렬을 구하시오.

$$f(x) = (3.9x_1 + x_2^3 - 2.3x_3^2, -2x_1x_2x_3^2)^T$$

12 다음 합성함수에 대해 답하시오.

$$f(x) = 2\left(\frac{1}{4}(1-2x)^2 - 1\right)^3 - 3\left(\frac{1}{4}(1-2x)^2 - 1\right)^2 - 3$$

- (1) 식 (2.53)에 따라 $i(x)$ 와 $h(x)$ 를 쓰시오.
- (2) 연쇄법칙을 이용하여 $f'(x)$ 를 구하시오.
- (3) $f'(0)$ 과 $f'(2.1)$ 을 계산하시오.

13 12번 문제의 $f(x)$ 에서 난수를 생성하여 초깃값 $x_0 = 2.1$ 을 얻었다고 가정하자. 식 (2.58)을 연속으로 적용하여 얻는 점 x_1, x_2, x_3 을 구하시오. 이때 학습률 $\rho = 0.1$ 을 사용하시오. 점점 낮은 곳을 향해 이동하는지 확인하시오.

14 다음 함수의 헤시안 행렬을 구하시오.

$$f(\mathbf{x}) = f(x_1, x_2, x_3) = x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 + x_1x_2x_3 + x_3^2$$

15 다음 함수에 대해 답하시오.

$$f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + 2x_2^2 - 4x_1 + 2x_2 - 24$$

- (1) 최소점과 최솟값을 분석적으로 구하시오.
- (2) 난수를 생성하여 초깃값 $\mathbf{x}_0 = (1.0, 0.9)^T$ 를 얻었다고 가정하고, 식 (2.58)을 연속으로 적용하여 얻는 점 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ 을 구하시오. 이때 학습률 $\rho = 0.1$ 을 사용하시오. (1)에서 구한 최소점을 향해 이동하는지 확인하시오.

16 [예제 2-11]의 테일러 급수에서 $\Delta = 0.2, 0.1, 0.01, 0.001$ 일 때 계산하고, Δ 가 정확도에 미치는 영향을 설명하시오.

17 [알고리즘 2-5]를 그 아래 코드로 바꾸어 쓰면, 한 번도 뽑히지 않는 샘플이 있을 수 있다. 그래도 문제가 없나? n 개의 샘플이 있고, 뽑는 일을 T 번 반복했을 때 한 번도 뽑히지 않은 샘플이 k 개일 확률을 쓰시오.

Hint 이항 분포를 적용하시오.

연습문제

1 [그림 3-10(b)]의 다층 퍼셉트론에 [그림 3-8(a)]에 있는 샘플 4개를 입력하여 예측 결과를 계산하시오. 모두 맞는지 확인하시오.

2 NOR 게이트와 AND 게이트의 동작을 데이터로 간주하면 다음과 같다. 이들을 100% 옳게 분류하는 퍼셉트론을 각각 제시하시오.

$$\text{NOR 분류} \begin{cases} x_1 = (0,0)^T, y_1 = 1 \\ x_2 = (1,0)^T, y_2 = -1 \\ x_3 = (0,1)^T, y_3 = -1 \\ x_4 = (1,1)^T, y_4 = -1 \end{cases} \quad \text{AND 분류} \begin{cases} x_1 = (0,0)^T, y_1 = -1 \\ x_2 = (1,0)^T, y_2 = -1 \\ x_3 = (0,1)^T, y_3 = -1 \\ x_4 = (1,1)^T, y_4 = 1 \end{cases}$$

3 3.4절의 오류 역전파 대신 식 (2.51)을 이용하여 그레이디언트를 근사 계산할 수 있다. 식 (2.51)을 사용하는 방법의 계산 효율을 분석하시오.

4 [알고리즘 3-5]를 배치 버전으로 바꾸어 쓰시오.

5 [표 3-1]에 있는 로지스틱 시그모이드의 1차 도함수를 그리시오. 이때 매개변수 s 의 범위는 $[-10, 10]$ 으로 하고 -10 에서 시작하여 1씩 증가시키면서 도함숫값을 계산하여 그래프를 그리시오. 값이 가장 큰 점과 가장 작은 점은 어디인지 확인하시오.

6 2장의 [그림 2-3]에 있는 퍼셉트론은 바이어스 노드가 없는 대신 임곗값 T 를 가진다. [그림 2-3]의 퍼셉트론은 [그림 3-3]의 퍼셉트론과 같은지 답하시오. 답에 대한 이유를 설명하시오.

7 [그림 3-4(b)]의 퍼셉트론에 대해 답하시오.

(1) w_1 과 w_2 의 값을 고정했을 때 w_0 이 어느 범위일 때 여전히 100% 옳게 분류하는가?

(2) w_0 과 w_2 의 값을 고정했을 때 w_1 이 어느 범위일 때 여전히 100% 옳게 분류하는가?

- 8 다음 훈련집합을 3차원 공간에 그리시오, 이 데이터를 최소 오류율로 분류하는 퍼셉트론을 제시하시오,

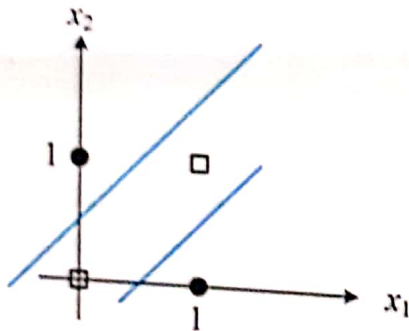
$$\begin{aligned} \mathbf{x}_1 &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ y_1 &= 1, \quad y_2 = -1, \quad y_3 = 1, \quad y_4 = -1, \quad y_5 = 1, \quad y_6 = -1 \end{aligned}$$

- 9 식 (3.7)에 있는 퍼셉트론의 목적함수를 다음과 같이 다르게 정의할 수 있다. 이 식을 미분하는 과정을 보이고, 미분 결과를 사용하여 가중치 갱신 규칙을 식 (3.9)처럼 제시하시오.

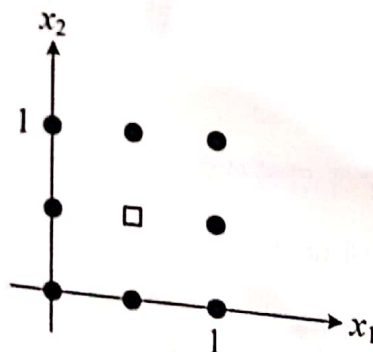
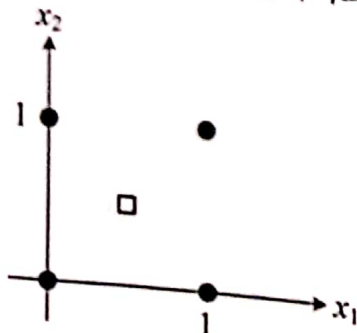
$$J(\mathbf{w}) = \sum_{l=1}^n (y_l - \tau(\mathbf{w}^T \mathbf{x}_l))^2$$

- 10 선형분리 불가능한 상황을 처리할 수 있도록 [알고리즘 3-1]을 수정하시오.

- 11 XOR 문제는 [그림 3-8(a)] 대신 다음 그림과 같이 해결할 수도 있다. 이 그림에 해당하는 다층 퍼셉트론을 제시하시오.



- 12 다음 분류 문제에 답하시오.



- (1) 이 분류 문제를 해결하는 다층 퍼셉트론을 각각 제시하시오.
- (2) 은닉 노드를 2개만 가진 퍼셉트론으로 오른쪽 상황을 해결할 수 있는지 답하시오. 가능하다면 은닉 노드 2개인 다층 퍼셉트론을 채택하는 것이 유리한지를 일반화 능력 측면에서 판단하시오.

Hint [그림 11-6]의 여백 개념을 사용하시오.

13 [알고리즘 3-4]의 부연 설명에서 “ X 에서 임의로 샘플 하나를 뽑는” 랜덤 샘플링 버전을 소개하였다. 랜덤 샘플링 버전이 MNIST 훈련집합을 사용한다고 가정하자. MNIST는 숫자 부류별로 6천 샘플을 가진다. [알고리즘 3-4]의 라인 2의 루프를 60만 번 반복한다고 가정한다. 10세대를 수행하는 셈이며 각 샘플은 평균 10번 선택된다. 다음 질문에 답하시오.

- (1) 어떤 샘플이 한 번도 선택되지 못할 확률은?
- (2) 어떤 부류에 속한 모든 샘플이 한 번도 선택되지 못할 확률은?
- (3) (1)과 (2)의 분석 결과를 토대로 랜덤 샘플링 방법의 유효성에 대한 의견을 제시하시오.

14 2차원 특징 공간을 p 개 직선으로 나누면 최대 $1 + \sum_{i=1}^p i$ 개의 영역으로 나뉜다. 주석 3을 참조하시오. 3차원 특징 공간을 p 개 평면으로 나누면 최대 몇 개 영역으로 나뉘는지 쓰시오.

15 현대 기계 학습은 [알고리즘 3-6]의 미니배치 스토캐스틱 경사 하강법SGD을 아주 널리 사용한다. 따라서 SGD는 사실상 표준으로 자리잡았다. 3장~5장이 참조한 논문 중에서 3개를 선정하여 미니배치 크기를 얼마로 설정했는지를 논문 제목과 함께 제시하시오.

16 [알고리즘 3-5]는 [알고리즘 3-4]의 행렬 표기 버전이다. 다음 두 라인에 대해 바꿔 쓸 수 있는 이유를 설명하시오.

[알고리즘 3-4]의 라인 7 → [알고리즘 3-5]의 라인 7

[알고리즘 3-4]의 라인 10 → [알고리즘 3-5]의 라인 10

17 분류 문제를 푸는 k -NN(k -nearest neighbor) 알고리즘이 있다. 이 알고리즘은 테스트 샘플에 가장 가까운 k 개의 샘플을 훈련집합에서 찾은 다음, k 개의 샘플의 부류를 보고 가장 빈도가 높은 부류로 분류한다.

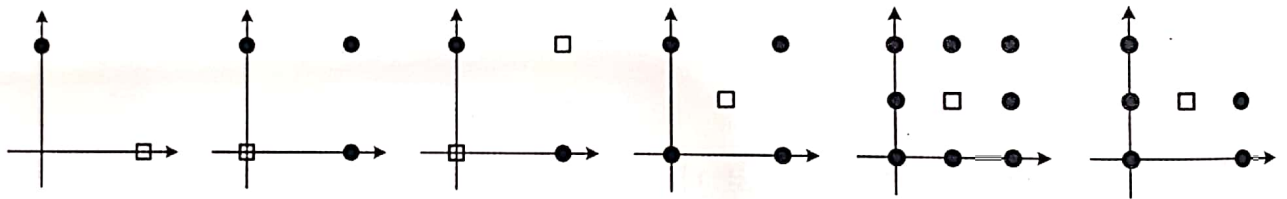
(1) k -NN을 이 책이 사용하고 있는 알고리즘 형태로 제시하시오.

(2) k -NN은 훈련집합을 모두 메모리에 저장하고 있어야 하는 메모리 기반 방법이다. MNIST를 훈련집합으로 사용할 때 메모리량을 산정하시오. (MNIST 훈련집합은 6만 개 샘플을 가지는데, 한 샘플은 784차원 특징 벡터로 표현된다.)

(3) 초당 1억 개의 사칙 연산을 수행하는 컴퓨터에서 5-NN으로 테스트 샘플 하나를 분류하는 데 걸리는 시간을 추정하시오.

(4) 다음 6가지 상황에 대해 1-NN이 공간을 어떻게 분할하는지 그리시오.

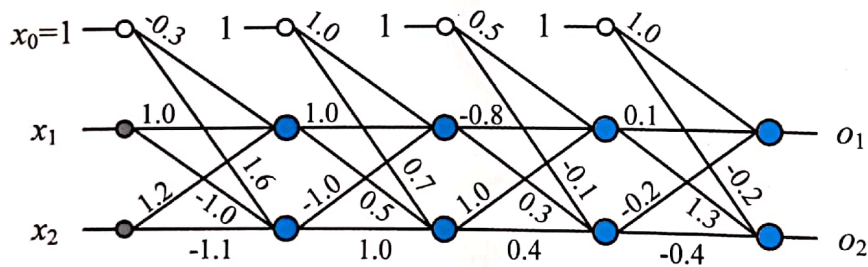
Hint 보로노이 도형(Voronoi diagram)으로 분할한다.



연습문제

- 1 [알고리즘 4-1]은 미니배치 버전이라는 점에서는 [알고리즘 3-6]을 따르지만 행렬 표기를 사용하지 않는 점은 [알고리즘 3-4]를 따른다. [알고리즘 4-1]을 행렬 표기로 바꾸어 쓰시오.
- 2 [그림 4-15]는 3차원 구조를 가진 데이터이다. [그림 4-15(a)]와 [그림 4-15(b)]를 위한 컨볼루션 수식을 식 (4.11)을 참조하여 각각 제시하시오.
- 3 [He2014]에서 제시한 가변 크기 데이터를 다루는 방법을 설명하시오.
- 4 컨볼루션 신경망의 아이디어는 1980년에 발표된 논문 [Fukushima1980]으로 알려져 있다. 한 동안 주목을 끌지 못하다가 1998년에 Proceedings of the IEEE 학술지에 발표된 [LeCun1998] 이후에 많은 연구가 이루어진다. 20여년간 주목을 끌지 못한 이유를 설명하시오.
- 5 다음은 은닉층이 3개인 DMLP이다.

Hint 계산은 Matlab 또는 Python을 사용하시오.



- (1) 가중치 행렬 U^1, U^2, U^3, U^4 를 식 (4.1)처럼 쓰시오.
- (2) $\mathbf{x} = (1, 0)^T$ 가 입력되었을 때 출력 \mathbf{o} 를 구하시오. 활성화함수로 로지스틱 시그모이드를 사용하시오.
- (3) $\mathbf{x} = (1, 0)^T$ 가 입력되었을 때 출력 \mathbf{o} 를 구하시오. 활성화함수로 ReLU를 사용하시오.
- (4) $\mathbf{x} = (1, 0)^T$ 의 기대 출력이 $\mathbf{o} = (0, 1)^T$ 일 때, 현재 1.0인 u_{12}^3 가중치를 0.9로 줄이면 오류에 어떤 영향을 미치는지 설명하시오.

- 6 [그림 4-31]은 ILSVRC의 물체 검출(object detection) 문제를 예시한다. 물체 검출 문제를 정의하고, 대회에 사용되는 성능 평가 기준을 설명하시오. 가장 최근 대회에서 달성한 성능을 조사하시오.
- 7 [그림 4-14]에서 나머지 8개 화소의 값을 계산하시오.
- 8 [그림 4-8(b)]에서 커널 $\begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ 을 적용한 결과를 쓰시오. 이때 0 덧대기를 하고 바이어스로 0.5를 사용하시오.
- 9 문제 8에서 보폭을 $s = 2$ 로 설정했을 때 컨볼루션 결과를 쓰시오.
- 10 문제 8의 결과에 최대 풀링과 평균 풀링을 적용한 결과를 각각 쓰시오. 보폭으로 1을 사용하시오.
- 11 [그림 4-14]를 [그림 4-15] 형식으로 다시 그리시오.
- 12 현재 시판되는 GPU 중에서 가격이 개인이 구입할 만한 제품 2종류를 제시하고, 사양을 구체적으로 기술하시오.
- 13 딥러닝을 리뷰한 Nature지(2015년 5월)에 실린 논문 [LeCun2015]를 자세히 읽고 1,000자 정도 분량의 요약 보고서를 쓰시오.

- 1 신경망의 출력이 $(0.4, 2.0, 0.001, 0.32)^T$ 일 때 softmax를 적용한 결과를 쓰시오.
- 2 softmax를 적용한 후 출력이 $(0.001, 0.9, 0.001, 0.098)^T$ 이고 레이블 정보가 $(0, 0, 0, 1)^T$ 일 때, 세 가지 목적함수, 평균제곱 오차, 교차 엔트로피, 로그우도를 계산하시오.
- 3 [예제 5-1]에서 $\lambda = 0.1$, $\lambda = 0.5$ 일 때를 계산하고 λ 에 따른 효과를 설명하시오. 이때 [그림 5-21]을 활용하시오.
- 4 훈련집합이 다음과 같다.

$$\mathbb{X} = \{x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, x_4 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}\},$$

$$\mathbb{Y} = \{y_1 = -1.8, y_2 = -3.6, y_3 = 4.2, y_4 = 7.2\}$$

- (1) [예제 5-1]처럼 리지 회귀를 적용하시오. $\lambda = 0.25$ 를 사용하시오.
 - (2) 테스트 샘플을 스스로 만들어 예측하시오.
 - (3) λ 를 반으로 줄일 때와 두 배로 키울 때 계산을 수행하고 λ 에 따른 효과를 분석하시오.
- 5 혈압, 키, 몸무게가 특징 벡터를 이룬다. 다음과 같이 훈련집합이 주어졌다.

$$\begin{pmatrix} 121 \\ 1.72 \\ 69.0 \end{pmatrix}, \begin{pmatrix} 140 \\ 1.62 \\ 63.2 \end{pmatrix}, \begin{pmatrix} 120 \\ 1.70 \\ 59.0 \end{pmatrix}, \begin{pmatrix} 131 \\ 1.80 \\ 82.0 \end{pmatrix}, \begin{pmatrix} 101 \\ 1.78 \\ 73.5 \end{pmatrix}$$

- (1) 퍼셉트론의 가중치 벡터가 $(-0.01, 0.5, -0.23)^T$ 이고 바이어스가 0이라고 했을 때, 훈련집합을 가지고 규모 문제를 설명하시오.
- (2) 식 (5.9)의 전처리를 적용한 후의 훈련집합을 쓰시오.
- (3) 전처리가 규모 문제를 완화하는지를 설명하시오.

6 VGGNet이 사용한 가중치 초기화 방법을 설명하시오.

Hint [Simonyan2016] 참조

7 ReLU 활성화함수의 발견은 딥러닝 발전에 큰 영향을 미쳤다. ReLU 발견 과정을 [Hanloser2000, Jarret2009, Nair2010, Glorot2011, Maas2013] 문헌을 순서대로 참고하여 요약 정리하시오.

Hint [Kuroikov2015] 참조

8 [알고리즘 5-2]는 식 (5.13)을 적용하였다. 식 (5.12)를 적용한 알고리즘을 쓰시오.

9 [알고리즘 5-5]의 Adam은 RMSProp에 식 (5.12)의 모멘텀을 적용하였다. RMSProp에 식 (5.13)의 네스테로프 모멘텀을 적용한 Nesterov_Adam 알고리즘을 작성하시오.

10 [예제 5-2]에서 w_6, w_7 을 계산하시오.

11 가중치 초기화는 $[-r, r]$ 범위의 난수를 생성해야 한다. 이 기능을 하는 C와 Python, Matlab 언어의 코드를 제시하시오.

12 5.5.1절의 임의 탐색이 로그 규모를 사용한다고 할 때, 로그 규모 난수 생성 기능을 하는 C와 Python, Matlab 언어의 코드를 제시하시오.

13 변수가 3개인 다음 함수에 [예제 5-2]와 같이 경사 하강법을 적용하시오.

$$J(\mathbf{w}) = J(w_1, w_2, w_3) = 2(w_1 - 2)^2 + (w_2 - 1)^2 + 2(w_3 - 1)^2$$

14 13번 문제에 있는 함수에 대해 [예제 5-3]과 같이 뉴턴 방법을 적용하시오.

15 [그림 5-22]에 대해 다음 주장의 진위 여부를 말하시오.

- θ 에서 출발하는 벡터 $-\rho \nabla J$ 는 $0 \sim 360$ 도 범위에서 어느 방향이라도 가질 수 있다.
- 벡터 $(-\rho \lambda, -\rho \lambda)^T$ 는 원점을 향한다.