# 2021-1 기계학습 중간고사 : 김휘용 교수님

**[문제 1] (m x n) 실수(real) 행렬 A와 B, 그리고 (n x 1) 실수 벡터 x와 b를 가정하라. A의 열벡터(column vector)들을 순서대로 a1, ..., an이라 하고, T는 transpose를 의미하며, '은 역행렬을 나타내고, 행렬 및 벡터의 곱을 .으로 표시한다고 하자. 다음 각 설명이 항상 참이면 O, 그렇지 않으면 X로 빈칸에 표시하시오. (6점: 각 문항당 0.5점)**

1. yT = xT.A 라 하면, y는 a1, ..., an 들의 선형결합이 된다.
2. D = diag(x) 라 하면, AD = DA가 성립한다.
3. D = diag(x) 라 하면, tr(D)는 x의 L1-norm과 같아진다.
4. D = diag(x) 라 하면, tr(DT.D)는 x의 L2-norm의 제곱과 같아진다.
5. A와 B가 대칭(symmetric) 정방행렬(square matrix)이면, A와 B는 non-singular가 된다.
6. A가 (n x n) 직교행렬(orthogonal matrix)이면, A의 Frobenius norm은 $\sqrt{n}$이 된다.
7. A와 B가 (n x n) 직교행렬이면, (A+B)' = A' + B' 가 성립한다.
8. m=10, n=20 일때, A의 column vector들 중 선형독립(linearly independent)인 벡터의 개수가 7개이면, A의 columspace의 차원은 7이 된다.
9. m=n=2 이고, a1 = [1 2]T, a2 = [2 1]T일 때, A의 고유값(eigenvalue) 중 가장 작은 값은 -1이 되며, 이 고유값에 해당하는 고유벡터(eigenvector)의 원소의 합은 0이 된다.
10. 위 9번 문항의 행렬 A는 positive definite이다.
11. AT.A = Q.R.QT 형태의 eigen-decomposition이 가능하며, Q는 A의 고유벡터들을 열벡터로 갖는 직교행렬이며, R은 A의 고유값들로 구성된 대각행렬이다.
12. u = bT.x, v = AT.x, w = xT.A.x 라 할 때, u, v, w의 gradient를 순서대로 나열하면 b, A, 2x 가 된다.

**[문제 2] 다음 표는 방의 개수와 집의 평수를 바탕으로 집값을 예측하는 문제에 대한 훈련데이터셋을 나타낸다. (이 표에서 x1은 방의 개수, x2는 평수, y는 집의 가격(억원)을 나타낸다. (6점: 각 문항당 1점)**

| x1 | x2 | y |
|---|---|---|
| 1 | 20 | 1 |
| 2 | 30 | 2 |
| 3 | 50 | 3 |
| 4 | 60 | 4 |

이 문제를 선형회기로 풀기 위해 다음과 같이 수학적 기호를 정의하였다. 물음에 답하시오.

- x(i): 표의 i번째 훈련예제의 특징(feature) 벡터
- y(i): 표의 i번째 훈련예제의 정답(target)
- h(x): 특징 벡터 x를 입력하였을 때 선형회귀 모델의 출력
- w: 선형회귀 모델의 파라메터(가중치) 벡터
- T: transpose를 나타내는 기호

1. Design matrix X의 두 번째 행의 값을 순서대로 나열하시오.
2. w = [-1 1 0.1]T 라 할 때, h(x(1)) 값을 구하시오.
3. w = [-1 0 0.1]T 라 할 때, 훈련집합 전체에 대한 MSE cost J(w)의 값을 구하시오.
4. w = [-1 0 0.1]T 라 할 때, 훈련집합 전체에 대한 MSE cost J(w)의 gradient 벡터를 구한 후, 그 원소값을 순서대로 나열하시오. (각 값은 분수로 기재하시오).

-3, -7, -109.9

5. 위 4번 문항의 w값과 gradient값을 이용하여 w를 업데이트 하고자 한다(Batch-mode Gradient Descent). 학습률을 4.0라 할 때, 업데이트된 파라메터 벡터를 구한 후, 그 원소값을 순서대로 나열하시오.

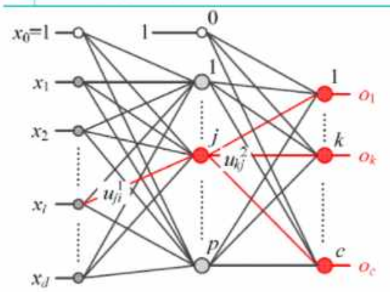
6. Feature normalization을 이용하여 Gradient Descent 알고리즘의 수렴 속도를 높이고자 한다. Feature normalization 후의 특징 벡터를 x'이라 할 때, x1에 대해서는 mean normalization만을 적용하여 $-1.5 \le x1' \le 1.5$가 되도록 하고, x2에 대해서는 mean normalization과 feature scaling을 모두 적용하여 $-0.5 \le x2' \le 0.5$ 가 되도록 하려 한다. x(3)에 대한 feature normalization 후의 특징백터 x(3)'의 원소의 값을 순서대로 나열하시오.

$\frac{60-40}{40}$ 0.5, 0.25

**[문제 3] 다음은 2층 퍼셉트론 학습을 위한 stochastic gradient descent 알고리즘을 나타낸다.**
**빈칸 [ 1 ]과 [ 2 ]에 알맞은 수식을 답안에 적으시오. (서술형 답안의 경우, 수식편집기를 이용할 수 있습니다.)**

**알고리즘3-4 다층 퍼셉트론 학습을 위한 스토캐스틱 경사 하강법**
**입력:** 훈련집합 $\mathbb{X}$와 $\mathbb{Y}$, 학습률 $\rho$
**출력:** 가중치 행렬 $\mathbf{U}^1$과 $\mathbf{U}^2$

```
U¹과 U²를 초기화한다.
repeat
  X의 순서를 섞는다.
  for (X의 샘플 각각에 대해)
     현재 처리하는 샘플을 x = (x_0, x_1, x_2, …, x_d)^T, y = (y_1, y_2, …, y_c)^T라 표기한다.
     x_0과 z_0을 1로 설정한다.  // 바이어스
                       // 전방 계산
     for (j=1 to p) zsum_j = u_j^1 x, z_j = τ(zsum_j)   // 식 (3.13)
     for (k=1 to c) osum_k = u_k^2 z, o_k = τ(osum_k)   // 식 (3.14)
                       // 오류 역전파
     for (k=1 to c) δ_k = [ 1 ]                          // 식 (3.22)
     for (k=1 to c) for (j=0 to p) Δu_kj^2 = -δ_k z_j    // 식 (3.23)
     for (j=1 to p) η_j = [ 2 ]                          // 식 (3.24)
     for (j=1 to p) for (i=0 to d) Δu_ji^1 = -η_j x_i    // 식 (3.25)
                       // 가중치 갱신
     for (k=1 to c) for (j=0 to p) u_kj^2 = u_kj^2 - ρΔu_kj^2   // 식 (3.21)
     for (j=1 to p) for (i=0 to d) u_ji^1 = u_ji^1 - ρΔu_ji^1   // 식 (3.21)
until (멈춤 조건)
```




**[문제 4] 다음 표는 방의 개수와 집의 평수를 바탕으로 종부세(종합부동산세) 납부 대상 여부를 판단하는 문제에 대한 훈련데이터셋을 나타낸다. 이 표에서 x1은 방의 개수, x2는 평수, y는 종부세 납부 대상 여부(1: 대상, 0: 비대상)을 나타낸다. 이 문제를 퍼셉트론으로 풀기 위해 다음과 같이 수학적 기호를 정의하였다. 물음에 답하시오. (4점: 각 문항당 1점)**

$g(2)+g(4)+g(7)+g(9)-2$

| x1 | x2 | y |
|---|---|---|
| 1 | 20 | 0 |
| 2 | 30 | 0 |
| 3 | 50 | 1 |
| 4 | 60 | 1 |

$-1+1+2=2$
$-1+2+3=4$
$-1+3+5=7$
$-1+4+6=9$

$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 20 & 30 & 50 & 60 \end{bmatrix}$ $\begin{bmatrix} g(2) \\ g(4) \\ g(7)-1 \\ g(9)-1 \end{bmatrix}$

- x(i): 표의 i번째 훈련예제의 특징(feature) 벡터
- y(i): 표의 i번째 훈련예제의 정답(target)
- h(x): 특징 벡터 x를 입력하였을 때 퍼셉트론 모델의 출력
- t(s): 계단 활성함수. t(s) = 1 if s>0, t(s) = -1 if s<0.
- g(s): 시그모이드 활성함수. g(s) = 1/(1+exp(-s)).
- w: 퍼셉트론의 파라메터(가중치) 벡터
- T: transpose를 나타내는 기호

$1-h_\theta(x)$

$\frac{exp(-s)}{1+exp(-s)} = \frac{1}{exp(s)+1}$

1. 활성함수로 t(s)를 사용하고, w = [-1 1 0.1]T 라 할 때, h(x(1)) 값을 구하시오. $-1+1+2=2$ 1
2. 활성함수로 g(s)를 사용하고, w = [-1 1 0.1]T 라 할 때, 표의 첫 번째 훈련예제에 대한 Cross-Entropy Loss J(w)를 구하여 간단히 하면 J(w) = log(**빈칸**) + 2 가 된다. **빈칸**에 알맞은 값을 적되, exp() 함수의 값은 계산하지 마시고 그대로 포함하시오. $=\log(exp(2)+1)(exp(4)+1)(exp(-7)+1)(exp(-9)+1)$
3. 활성함수로 g(s)를 사용하고, w = [-1 1 0.1]T 라 할 때, 표의 첫 번째 훈련예제에 대한 Cross-Entropy Loss J(w)의 gradient 벡터 ∇J(w)를 구하여 간단히 하면 ∇J(w) = g(2)·[**빈칸**]T 가 된다. **빈칸**에 알맞은 원소값들을 적으시오. $x^T(g(xw)-y)$
4. 활성함수로 g(s)를 사용하고, w = [a b c]T 라 할 때, 결정경계(decision boundary) d(x1,x2)의 식을 적으시오. d(x1,x2) =

$g(xw)$

**[문제 5] 기계학습 분류 문제에서 훈련집합에 대해 MSE cost function J(w)를 전역최소(global minimum)로 하는 최적 파라메터 벡터 w*를 구했다고 가정하자. 이때 사용한 기계학습 모델을 h(w)라 할 때, 다음 각 항목의 설명에 대해 항상 맞으면 O로, 그렇지 않으면 X로 빈칸을 채우시오. (둘다 맞아야 1점)**

1. h(w*) = 0 이다.
2. J(w*) = 0 이다.

**[문제 6] 다음의 부등식 조건부 최적화 문제(inequality constrained optimization problem)에 대해, 다음 물음에 답하시오. (서술형 답안은 수식편집기를 이용할 수 있습니다.) (4점)**

- Minimize J(x,y) = $2x^2+y^2$
- Subject to $x+2y \geq 1$

1. 위 최적화 문제에 대한 라그랑지(Lagrange) 함수를 기재하시오. (1점)
2. 위 최적화 문제에 대한 KKT 조건 4가지를 기재하시오. (1점)
3. 위 최적화 문제를 풀어 최적해 (x*, y*)을 구하고, 해가 하나만 존재하는 이유를 KKT 조건을 활용하여 설명하시오. (2점)

**[문제 7] 다음 각 문제 1~4에 대해 가장 적합한 학습 알고리즘을 A~C 중에서 골라 빈 칸에 넣으시오. (다 맞아야 1점)**

1. 심장병 환자들 10,000명의 30년간 의료기록을 바탕으로 나이에 따른 재발 가능성을 예측하는 문제: B
2. 이메일의 내용을 바탕으로 스팸인지 아닌지 구분하는 문제: A
3. 코로나 백신 접종자 10,000명의 접종후 신체반응 데이터를 바탕으로 신체반응이 몇 가지 패턴으로 나뉘는 지 찾아내는 문제: C

A. 분류(classification)를 위한 지도학습(supervised learning) 알고리즘
B. ~~회귀(regression)를 위한 지도학습(supervised learning) 알고리즘~~
C. 군집화(clustering)을 위한 비지도학습 알고리즘

**[문제 8] 아래와 같이 주어진 분류(classification) 문제를 해결하는 다층퍼셉트론(MLP)를 설계하시오.**
**설계에 필요한 기호는 설명에 필요한 만큼 정의하여 사용하되, 아래 미리 정의된 기호 및 내용은 따르도록 하시오. (본 문제에서는 아래 분류문제를 오류 없이 해결하면서도 사용된 뉴런의 개수를 최소화하는 설계를 좋은 설계라고 정의하며, 좋은 설계일 수록 높은 점수를 받도록 채점할 예정임.)**

-입력층 노드의 개수: d=3.
- 출력층 노드의 개수: c=? ~~2~~ ~~2~~ 1
- 사용된 층의 개수: M=? 2
  (이때, 층의 개수는 은닉층과 출력층을 모두 포함한다.)
- m번째 층의 노드(뉴런)의 개수: p(m)=? $p(1)=3$, $p(2)=1$
- m번째 층의 가중치 행렬: Um=?
  (이때, m번째 층의 j번째 노드(뉴런)에 대해 i번째 입력과 연결된 가중치는 Um(j,i)로 정의하며, Um은 Um(j,i)를 원소로 갖는 행렬이다.)
- t(s): 계단 활성함수. t(s) = 1 if s>0, t(s) = -1 if s<0.
- g(s): 시그모이드 활성함수. g(s) = 1/(1+exp(-s)).
- r(s): relu 활성함수. r(s) = max(0, s).

$$U_m = \begin{bmatrix} U_m(0,0) & \cdots & U_m(p(m),0) \\ U_m(0,1) & & \vdots \\ \vdots & & \\ U_m(0,p(m-1)) & & U_m(p(m), \;) \end{bmatrix}$$

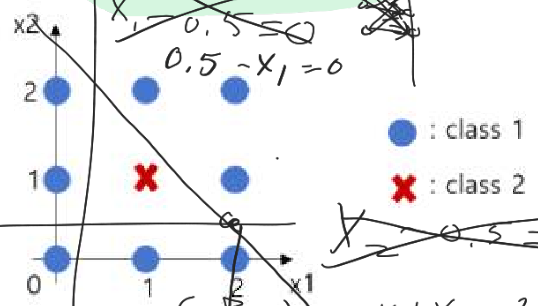
$U_m(j,0)$

$U_m(j,i) \to O_j$

$$U_1 = \begin{bmatrix} U_1(1,0) & U_1(2,0) & U_1(3,0) \\ U_1(1,1) & U_1(2,1) & U_1(3,1) \\ U_1(1,2) & U_1(2,2) & U_1(3,2) \end{bmatrix}$$

~~$x_1 - 0.5 = 0$~~

$0.5 - x_1 = 0$




~~$x_2 - 0.5 = 0$~~ $0.5 - x_2 = 0$

$(2, 0.5)$

$x_1 + x_2 - 2.5 = 0$

1. 본인 설계의 핵심 아이디어를 설명하시오.
2. 위 정의된 변수들 중 "?"로 표시된 변수(또는 수식, 또는 행렬)의 값을 설계 결과를 토대로 답안에 제시하시오.
3. 설계 결과로 위 분류문제를 오류 없이 풀 수 있음을 증명하시오.

영역을 7개로 나누어

위 그림을 보면 직선 4개로 class1과 class2를 오류 없이 분류해 낼 수 있다. 따라서 은닉층을 하나 더 만들어서 해당 은닉층의 노드 수가 3개가 되도록 한다.

정답 : 정답 안알려준 것도 있음

[문제 1] X X X O X O O O O X O O

[문제 2] (2, 30) , 1, (1/4), (1/2, 7/4, 55/2), (-3, -7, -109.9), (1/2, 1/4)

[문제 4] 1, (1+exp(-2)), 3번 답 모름, a+b.x1+c.x2=0   또는 a+bx1+cx2=0

[문제 5] X, X

[문제 7] B A C