

Final Exam

[SCS4049-2] Machine Learning and Data Science

Dec. 14, 2020

- 시험시간: 10시 00분 – 11시 50분, 연장 없음 (아래 표 참조)
- 만점: 100점 + 보너스 5점
- 문제지는 표지 포함 총 5쪽입니다.
- 객관식 또는 단답형이 표시되지 않은 경우 서술형입니다.
- 객관식과 단답형은 부분점수가 없습니다.
- 기입 사항
 - 답안지 첫번째 쪽: 학번, 학과, 이름, 답안지 전체 쪽수
 - 답안지 각 쪽마다: 해당 답안지의 쪽수
- 교수와 조교의 지시에 따라주세요.
 - 카메라와 마이크는 항상 켜주세요.
 - 지시가 있기 전까지 미팅룸을 나가지 마세요.

시간	해야할 일	비고
시험 시작 전부터 10시 00분까지	카메라 설정 완료 미팅룸 입장 완료	준비에 도움이 필요하면 사전에 담당 교수에게 문의
10시 00분부터 카메라 설정 확인시까지	카메라 설정 확인	수강생 본인 책임 하에 준비되지 않은 사람은 시험 자격 박탈 최대 10시 10분을 넘기지 않으며 그때까지 협조하지 않는 경우 자격 박탈
카메라 설정 확인시부터 11시 35분까지	시험 응시	
11시 35분부터 11시 40분까지	답안을 카메라에 차례로 보여줌	시험 일괄 종료
11시 40분부터 11시 50분 59초까지	답안을 스캔하여 메일 전송	메일 도착 시간 기준 11시 51분 00초부터 0점 처리

객관식

답안지에 아래와 같은 표를 그려서 답해주세요.

문제 1번	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
정답 (T/F)								
문제 2번	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
정답 (T/F)								

1. 다음에 대해 (T)rue 또는 (F)alse로 대답하세요. [각 1.5점 / 총 12점]

- (a) Covariance matrix는 행과 열의 개수가 동일한 정사각행렬(square matrix)이다.
- (b) Neural network의 cost function으로 자주 사용되는 cross entropy $E = -\sum t_j \log y_j$ 는 작을수록 좋다.
- (c) 구체적으로 프로그래밍하지 않아도 데이터로부터 스스로 학습하여 특정 임무를 수행할 수 있는 능력을 컴퓨터가 갖도록 구현하는 AI의 한 분야를 머신러닝이라고 한다.
- (d) PCA를 이용한 dimensional reduction에서는 singular value가 큰 순서대로 선택해야 정보의 손실을 최소화할 수 있다.
- (e) 스칼라를 벡터로 미분하면 벡터가 나온다.
- (f) Neural network를 classification 문제에 적용하기 위해서 출력층의 unit 숫자는 전체 class의 개수와 동일해야 한다.
- (g) Convolutional neural network도 sequence 입력을 다룰 수 있다.
- (h) Recurrent neural network는 입력과 출력의 sequence 길이가 다를 수 있다.

2. 다음에 대해 (T)rue 또는 (F)alse로 대답하세요. [각 2.5점 / 총 20점]

- (a) 주어진 A 행렬에 SVD를 적용해 얻어진 $A = U\Sigma V^T$ 에 대하여, $U^T U = I$ 와 $V^T V = I$ 가 성립한다.
- (b) Stochastic gradient descent는 무작위로 선택한 한 개의 sample에 대해서 gradient를 계산한다.
- (c) Convolutional neural network에서 maxpooling layer를 거치면 이미지의 크기가 줄어든다.
- (d) Convolutional neural network에서 convolution layer에 stride를 적용하면 이미지의 크기가 줄어든다.
- (e) Linear regression에서 RMSE, MSE, SSE의 해는 모두 동일하다.
- (f) Neural network의 activation function 중 하나인 rectified linear unit (ReLU)은 다음과 같이 표현할 수 있다.

$$\text{ReLU}(z) = \max(0, z) \tag{1}$$

- (g) Soft SVM에서 margin의 침범 여부와 그 정도는 slack variable로 표현된다.
- (h) 어떤 확률분포가 갖는 정보량을 entropy를 통해 측정할 수 있다.

주관식

- (단답형) MoG clustering 알고리즘에서 responsibility의 정의를 쓰세요. [3점]
- (단답형) 주어진 행렬 A 를 이용하여 2차원 상에서 반지름이 1인 원 위의 벡터들의 집합 $\{x \mid x^T x = 1\} \subset \mathbb{R}^2$ 을 변환하면 타원의 형태로 나타납니다. 그 타원의 장축과 단축의 방향과 크기를 행렬 A 에 SVD를 적용한 결과 $A = U\Sigma V^T$ 로 나타내세요. [9점]
- (단답형) 아래 그림의 python script를 실행시켰을 때 출력되는 각 layer별 output의 shape를 쓰세요. 즉, 빈 칸에 들어갈 여섯개의 숫자를 차례대로 대답하세요. 단, stride의 default 값은 1이며, padding은 적용하지 않았습니다. [6점]

```
In [2]: model = models.Sequential()
model.add(layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)))
model.add(layers.Conv2D(64, (2, 2), activation='relu'))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))
```

```
In [3]: model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, ①, ②, 32)	320
conv2d_1 (Conv2D)	(None, ③, ④, 64)	8256
conv2d_2 (Conv2D)	(None, ⑤, ⑥, 64)	36928
Total params: 45,504		
Trainable params: 45,504		
Non-trainable params: 0		

- (단답형) 아래 그림의 python script는 무작위로 생성한 샘플 X 의 covariance matrix를 구하는 과정입니다. 가려진 위치에 들어갈 함수의 이름과 그 함수에 입력되는 두번째, 세번째 인자를 쓰세요. [4점]

```
In [5]: 1 import numpy as np
2 import numpy.matlib
3 import matplotlib.pyplot as plt
4
5 N = 1000
6 D = 2
7
8 X = np.random.randn(D, N)
9 A = np.random.randn(D, D)
10 X = A.dot(X)
11
12 mu = np.mean(X, axis = 1).reshape(D, 1)
13 C = X - np.matlib. 함수? (mu, ?, ?)
14 C = (1/N)*C.dot(C.T)
15
```

7. 다음의 학습 데이터를 이용한 linear regression에 대해 다음의 질문에 답하세요.

sample index (i)	1	2	3	4	5
input $x^{(i)}$	1	1	2	2	3
output $y^{(i)}$	4	5	7	8	10

- (a) (단답형) Linear regression에서 파라미터 θ 를 찾을 때, 주로 최소화하고자 하는 값으로 주어지는 식 (2)의 함수 이름 전체를 영어로 쓰세요. [2점]

$$\sqrt{\frac{1}{m} \sum_{i=1}^m (\theta^T \mathbf{x}^{(i)} - y^{(i)})^2} \quad (2)$$

- (b) (단답형) Bias를 포함한 1차 polynomial에 대하여 주어진 학습 데이터의 design matrix \mathbf{X} 를 구하세요. [3점]

- (c) (단답형) Bias를 포함한 1차 polynomial에 대해 식 (2)를 최소화하는 해 $\hat{\theta}$ 를 구하세요. [3점]

8. (단답형) Neural network의 backpropagation은 다음의 여섯단계로 요약할 수 있습니다. 주어진 notation을 사용하여 각 단계의 빈칸 (a)–(e)를 완성하세요. [각 2점 / 총 10점]

- **Input:** 주어진 입력 \mathbf{x} 로부터 입력층의 activation \mathbf{a}^1 를 계산
- **Feedforward:** 각각의 층 $l = 2, 3, \dots, L$ 에서 다음을 계산

$$\mathbf{z}^l = \mathbf{W}^l \boxed{(a)} \quad \mathbf{a}^l = \sigma^l(\mathbf{z}^l) \quad (3)$$

- **Output error:** 출력층에서 cross entropy를 이용해 local gradient를 계산

$$\delta^L = \boxed{(b)} \quad (4)$$

- **Backpropagate error:** 출력층부터 거꾸로 다음의 식을 계산

$$\delta_k^l = \boxed{(c)} \sum_{j=1}^{M_{l+1}} \boxed{(d)} \quad (5)$$

- **Output gradient:** 각 weight에 대해 gradient 계산

$$\boxed{(e)} = \delta_k^l a_j^{l-1} \quad (6)$$

- **Update weight:**

$$w_{kj}^l \leftarrow w_{kj}^l - \eta \delta_k^l a_j^{l-1} \quad (7)$$

9. (단답형) 다음 두 행렬을 convolution한 결과를 구하세요. 단, zero padding을 적용합니다. [5점]

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 9 \\ 1 & 1 & 2 & 7 & 8 \\ 1 & 2 & 8 & 8 & 1 \\ 2 & 9 & 7 & 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix} = ? \quad (8)$$

10. (단답형) 다음 행렬에 2×2 의 maxpooling을 적용한 결과를 구하세요. 단, stride는 1×2 를 적용합니다. [3점]

$$\begin{bmatrix} 2 & 5 & 9 & 5 & 1 & 2 \\ 1 & 8 & 8 & 3 & 2 & 4 \\ 7 & 2 & 6 & 9 & 2 & 3 \\ 9 & 1 & 5 & 8 & 3 & 6 \end{bmatrix} \quad (9)$$

11. Long short-term memory (LSTM)에 대한 다음의 질문에 답하세요.

- (a) 하나의 gate는 어떻게 구성되어 있으며, 어떻게 동작하는지 설명하세요. [3점]
- (b) 다음의 notation을 이용하여 LSTM cell의 구조를 그리세요. [5점]
 - 입력되는 값: cell state C_{t-1} , hidden state h_{t-1} , input x_t
 - 출력되는 값: cell state C_t , hidden state h_t
- (c) (b)에서 forget gate는 무엇이고 그 역할이 무엇인지 설명하세요. [2점]
- (d) (b)에서 input gate는 무엇이고 그 역할이 무엇인지 설명하세요. [2점]
- (e) (b)에서 hidden state는 어떻게 업데이트 되는지 설명하세요. [2점]

12. Variational autoencoder가 무엇인지 설명하세요. [6점]