

Mid-Term Exam

[SCS4049-2] Machine Learning and Data Science

Oct. 26, 2020

- 시험시간: 10시 00분 - 11시 50분, 연장 없음 (아래 표 참조)
- 만점: 100점
- 문제지는 표지 포함 총 4쪽입니다.
- 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.
 - 지시가 있기 전까지 키보드를 만지지 마세요.
- 답안지 실물 제출: 28일(수)까지 만해관 210호

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

1. 다음에 대해 (T)rue 또는 (F)alse로 대답하세요. [각 1점 / 총 10점]

- 구체적으로 프로그래밍하지 않아도 데이터로부터 스스로 학습하여 특정 임무를 수행할 수 있는 능력을 컴퓨터가 갖도록 구현하는 AI의 한 분야를 딥러닝이라고 한다.
- K-means 알고리즘과 MoG 알고리즘 모두 cluster의 갯수를 먼저 명시해주어야 한다.
- 강화학습(reinforcement learning)은 행동(action)에 대한 보상(reward)을 필요로 한다.
- Clustering은 대표적인 unsupervised method이다.
- Regression과 classification은 target output이 연속인지 아닌지에 따라 구분된다.
- Mini-batch learning은 online learning과 batch learning의 절충안이다.
- 확률의 합은 항상 1이다.
- Decision tree에서 경계에 의해 데이터가 잘 나뉘어질수록 Gini impurity measure는 0에 가깝다.
- K-fold cross validation에서 k는 10으로 정해져있다.
- Mini-batch learning은 batch learning에 비해 컴퓨팅 자원을 더 많이 요구한다.

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

- Lagrangian multiplier는 항상 0보다 작아야한다.
- SVM에서 데이터 \mathbf{x}_n 으로부터 decision boundary까지의 거리는 $\frac{t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b)}{\|\mathbf{w}\|}$ 이다.
- Linear regression에서 RMSE, MSE, SSE의 해는 모두 동일하다.
- Hard SVM에서 margin의 침범 여부와 그 정도는 slack variable로 표현된다.
- 과적합에 대응하기 위해 모델의 복잡도에 제한을 두는 것을 정규화(regularization)라고 한다.
- Linear discriminant function의 decision boundary는 \mathbf{w} 에 평행하다.
- SVM은 convex optimization problem에 속한다.
- 모델의 복잡도가 필요 이상으로 너무 높아 일정 수준 오차를 허용하지 않고 모든 sample에 대해 대응하기 시작할 때 미적합(underfitting)이 일어날 수 있다.
- Gradient descent는 무작위로 선택한 한 개의 sample에 대해서 gradient를 계산한다.
- Soft voting classification은 decision에 대한 확률값을 갖는 classifier에만 적용할 수 있다.

주관식

3. 다음의 확률 질량 함수(probability mass function)를 따르는 확률변수 X 에 대해 다음의 질문에 대해 답하세요. [5점]

random variable X	1	2	3	4	5
probability $\Pr(X = x)$	0.1	0.1	0.5	0.2	0.1

- (a) (단답형) 평균 $E(X)$ 와 분산 $\text{Var}(X)$ 를 구하세요. [3점]
 (b) (단답형) 확률 분포 함수 $F_X(x) = \Pr(X \leq x)$ 를 직접 그리세요. [2점]
4. 다음의 학습 데이터를 이용한 linear regression에 대해 다음의 질문에 답하세요. [10점]

sample index (i)	1	2	3	4	5	6	7	8	9	10
input $x^{(i)}$	1	1	2	2	3	4	4	5	6	7
output $y^{(i)}$	4	5	7	8	10	12	13	15	19	21

- (a) (단답형) Linear regression에서 파라미터 θ 를 찾을 때, 주로 최소화하고자 하는 값으로 주어지는 식 (1)의 함수 이름 전체를 영어로 쓰세요. [2점]

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

- (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 (1)를 최소화하는 해 $\hat{\theta}$ 를 구하세요. 단, 근사값은 식 (2)을 활용하고 소수점은 세자리까지만 구하세요. [5점]

$$\frac{10}{385} \approx 0.026, \quad \frac{35}{385} \approx 0.091, \quad \frac{161}{385} \approx 0.418 \quad (2)$$

5. K-means clustering에 대한 다음의 문제에 대해 답하세요. [10점]

- (a) (단답형) K-means clustering의 결과로 나오는 decision boundary의 형태를 부르는 용어를 쓰세요. [3점]
 (b) K-means clustering 알고리즘에서 iteration을 통해 반복되는 두 과정에 대해 설명하세요. 그림으로 설명해도 되고 수식으로 설명해도 됩니다. [7점]

6. (단답형) MoG clustering 알고리즘에서 responsibility의 정의를 쓰세요. [10점]

7. Support vector machine (SVM)에 대한 다음의 문제에 대해 답하세요. [10점]
- (a) (단답형) Convex optimization 문제를 변환한 Lagrangian dual problem의 해가 반드시 만족하는 아래의 4가지 조건을 무엇이라고 하나요. 줄인 말로 쓰셔도 됩니다. [5점]
- i. primal constraints: $f_i(x) \leq 0, h_i(x) = 0$
 - ii. dual constraints: $\lambda \geq 0$
 - iii. complementary slackness: $\lambda_i f_i(x) = 0$
 - iv. gradient of Lagrangian: $\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$
- (b) (단답형) 위의 4가지 조건 중에서 support vector 여부를 결정하는데 가장 중요한 조건은 무엇인지 그 번호나 이름을 쓰세요. [5점]
8. Hard voting classifier가 ensemble의 여러가지 결과에 대해서 어떻게 하나의 결과를 선택하는지 설명하세요. [5점]
9. (단답형) 주어진 학습 데이터 $\mathbf{x}^{(i)}$ ($i = 1, 2, \dots, N$)와 임의의 함수 $k(\mathbf{x}, \mathbf{x}')$ 가 있을 때, Gram matrix의 i 행 j 열 G_{ij} 는 어떻게 정의되는지 쓰세요. [5점]
10. (단답형) Binary classification의 confusion matrix에서 예측값(prediction)과 참값(ground truth)으로 구분되는 4가지 경우의 수에 대한 명칭을 표로 나타내세요. [3점]
11. (단답형) Bayes' theorem을 쓰고 prior, likelihood, posterior probability가 무엇에 해당하는지 각각 쓰세요. [2점]
12. Lagrangian multiplier를 이용하여 $3x^2 + 4y^2 = 1$ 를 만족하는 x, y 에 대해서 $f(x, y) = x + y$ 를 최소화하는 해를 구하세요. [10점]