기계학습원론 중간고사
2025 가을

강의자: 조재민

강의조교: 이재웅

이름: ______________________

학번: ______________________

2025년 10월 20일

문제 1. 명제의 참(O) 또는 거짓(X)을 판단하시오. (정답: +2점, 오답: -2점, 미기재: 0점)

1. Linear Regression은 닫힌 형태 솔루션(Closed-form solution)이 존재하나 Logistic Regression은 닫힌 형태 솔루션이 존재하지 않아 보통 Stochastic Gradient Desecent로 학습한다.
2. 연속 확률 변수 X의 확률 밀도 함수(Probability Density Function) $p(x)$는 모든 x에 대해 $0 \le p(x) \le 1$을 만족한다.
3. 임의의 $x, x' \in X$, $t \in [0,1]$에 대해 $f(tx + (1-t)x') \ge tf(x) + (1-t)f(x')$인 함수 $f : X \to \mathbb{R}$에서, 어떤 지점의 그래디언트가 0이면 항상 전역 최댓값이다.
4. 이산 확률 분포 P, Q에서 P가 K개의 클래스($x \in \{c_1, \ldots, c_K\}$)에 대한 균등 분포 ($P(x = c_i) = 1/K$)일 때, $KL(Q||P) = \log K - H(Q)$ 이다. (단, $H(Q)$는 분포 Q의 엔트로피이다.)
5. 총 3개의 클래스를 분류하는 소프트맥스 회귀(Softmax regression)에서 각 클래스에 대한 로짓이 21, 6, 3일 때, 소프트맥스 함수 적용 후 각 클래스에 대한 확률은 0.7, 0.2, 0.1이다. (단, 부동소수점 오차는 무시한다.)

문제 2. [**8점**] 다음은 머신러닝 모델의 학습에 대한 설명이다. 옳은 문장을 모두 고르시오.

a. 테스트 세트(Test set)는 주로 모델의 하이퍼파라미터를 탐색하기 위해 사용된다.
b. 데이터 증강(Data augmentation)은 기존 훈련 데이터를 변형하여 훈련 샘플의 수를 늘림으로써 과소적합(Underfitting)을 해결하는 데 주로 사용된다.
c. 에포크(Epoch) 수가 고정일 때, 경사 하강법(Gradient descent)의 학습률(Learning rate)을 늘릴수록 훈련 오차(Training error)는 작아진다.
d. 동일한 손실 함수(Loss function)에 대해 적절한 학습률 감소 조건을 적용하여 충분히 학습한다면, Stochastic Gradient Descent (SGD)와 Batch Gradient Descent은 항상 동일한 최적해에 도달한다.
e. Mini-batch SGD에서 데이터가 N개, 배치 크기(Batch size)가 B이면 한 에포크의 업데이트 횟수는 $\lceil N/B \rceil$이다. (단, N이 B로 나누어 떨어지지 않더라도 남은 데이터를 학습에 사용한다.)

문제 3. [**8점**] 다음 각 연산이 머신러닝 모델의 편향(Bias)과 분산(Variance)에 미치는 영향으로 가장 적절한 것을 고르시오.

1. 회귀 모델에 정규화(Regularization) 적용:
   i. 편향: ◯ 증가 ◯ 감소 ◯ 변화 없음
   ii. 분산: ◯ 증가 ◯ 감소 ◯ 변화 없음
2. 회귀 모델의 복잡도 증가:
   i. 편향: ◯ 증가 ◯ 감소 ◯ 변화 없음
   ii. 분산: ◯ 증가 ◯ 감소 ◯ 변화 없음
3. 훈련 데이터의 크기(Data size) 증가:
   i. 편향: ◯ 증가 ◯ 감소 ◯ 변화 없음
   ii. 분산: ◯ 증가 ◯ 감소 ◯ 변화 없음
4. k-NN(k-Nearest Neighbors) 모델의 이웃 수(k) 증가:
   i. 편향: ◯ 증가 ◯ 감소 ◯ 변화 없음
   ii. 분산: ◯ 증가 ◯ 감소 ◯ 변화 없음

문제 4. [**8점**] 아래 문제에 답하시오.

1. $f(\theta) = \frac{1}{\theta^2}\exp(-\frac{1}{2\theta^2})$를 θ에 대해 미분하라.
2. $\log f(\theta)$를 θ에 대해 미분하라.
3. $f(\theta)$는 어떤 지점에서 최댓값을 가지며, 그 최댓값은 얼마인가? (단, $\theta > 0$)

문제 5. [6점] 세 파라미터 w_0, w_1, w_2에 대한 손실 함수가 다음과 같다:

$$E(w_0, w_1, w_2) = w_0^4 - 2w_0^2 + w_1^2 + w_2^2 + 2025$$

초기 파라미터 값 $(w_0, w_1, w_2) = (0, 1, -1)$, 학습률 $\alpha = 0.5$일 때, 경사 하강법을 한 번 수행한 후의 파라미터 w_0, w_1, w_2의 값을 각각 구하여라.

문제 6. [8점] 다음과 같은 다중 클래스 분류 결과가 주어져 있다:

실제 레이블 (Ground-truth):	[A, B, C, A, B, B]
모델 예측 (Model prediction):	[B, C, C, A, B, A]

1. 클래스 B에 대한 Precision, Recall, 그리고 F1-score를 계산하여라.
2. 모델의 예측이 변경되어 Precision이 증가할 경우, Recall은 항상 감소하거나 같다. (O / X)

문제 7. [8점] 녹색, 하얀색, 노란색의 세 가지 색상의 송편이 있다. 녹색 송편은 팥소 2개, 깨소 3개, 콩소 1개가 있고, 하얀색 송편은 팥소 1개, 깨소 2개, 콩소 3개가 있으며, 노란색 송편은 팥소 3개, 깨소 1개, 콩소 2개가 있다. 송편을 $p(\text{녹색}) = 0.4$, $p(\text{하얀색}) = 0.4$, $p(\text{노란색}) = 0.2$의 확률로 무작위로 선택하고, 선택된 송편에서 하나의 소를 맛보려 한다. (단, 각 소는 동일한 확률로 선택된다.)

1. 깨소 송편을 선택할 확률은 얼마인가?
2. 선택한 송편이 깨소라는 것을 관찰했을 때, 이 송편이 녹색 송편이었을 확률은 얼마인가?

문제 8. [8점] 다음 데이터 샘플을 고려하자:

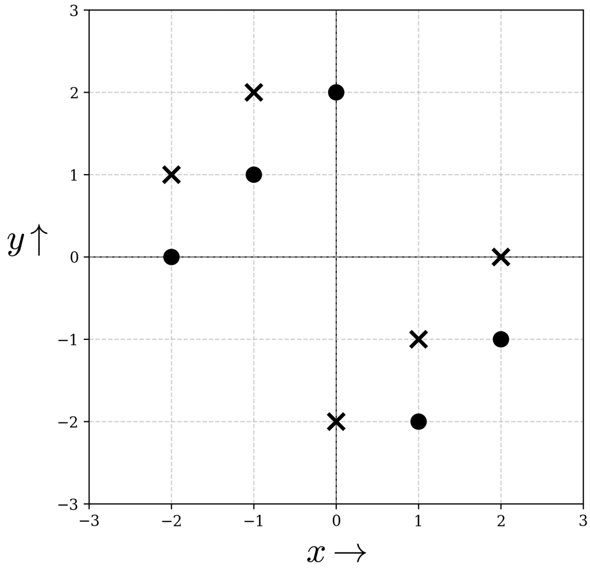
$$(x_1, y_1) = (1, 2), \quad (x_2, y_2) = (2, 5), \quad (x_3, y_3) = (3, 5)$$

선형 회귀 모델 $f(\mathbf{x}) = w_0 + w_1 x$, 전체 데이터셋 $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{n}$ $(n = 3)$에 대해, 손실 함수 $E(\mathbf{w})$는 다음과 같이 정의된다:

$$E(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - f(\mathbf{x}^{(i)}))^2$$

1. 정규방정식(Normal Equation)을 사용하여 손실 함수 $E(\mathbf{w})$를 최소화하는 파라미터 $\mathbf{w} = [w_0, w_1]^\top$를 구하고, 이를 통해 선형 회귀 직선 $y = w_0 + w_1 x$를 구하시오.
2. 학습 데이터셋 $\mathcal{D}$에 새로운 데이터 (x_4, y_4)를 추가하여($n = 4$) 파라미터 $\mathbf{w}$를 다시 구하였다고 하자. 회귀 직선이 변하지 않았다면, 이 데이터는 반드시 앞서 구한 회귀 직선 위에 존재한다. (O / X)

문제 9. [10점] 다음 2차원 데이터셋에 대해, 유클리드 거리(Euclidean distance)를 사용하는 k-NN(k-Nearest Neighbors) 분류를 수행하고자 한다. 그림의 '×'와 '•'는 서로 다른 두 클래스를 나타내며, 데이터는 총 $n = 10$개의 샘플로 구성되어 있다.



Leave-one-out Cross-Validation (LOOCV)는 전체 n개 샘플 중 하나를 검증용으로, 나머지 $n-1$개를 학습용으로 사용하는 과정을 모든 샘플에 대해 (총 n번) 반복하여 오류를 측정하는 평가 방식이다. LOOCV 오류는 다음과 같이 계산된다:

$$\text{LOOCV 오류} = \frac{\text{잘못 분류된 검증용 샘플의 총 수}}{\text{검증용 샘플의 총 수}}$$

예를 들어, 10개의 샘플 중 각 샘플에 대해 남은 9개의 샘플로 학습을 했을 때, 3개가 잘못 분류되었다면 모델의 LOOCV 오류는 0.3이다.

1. 다음 k-NN 모델에서의 LOOCV 오류를 계산하시오.
   (a) 1-NN

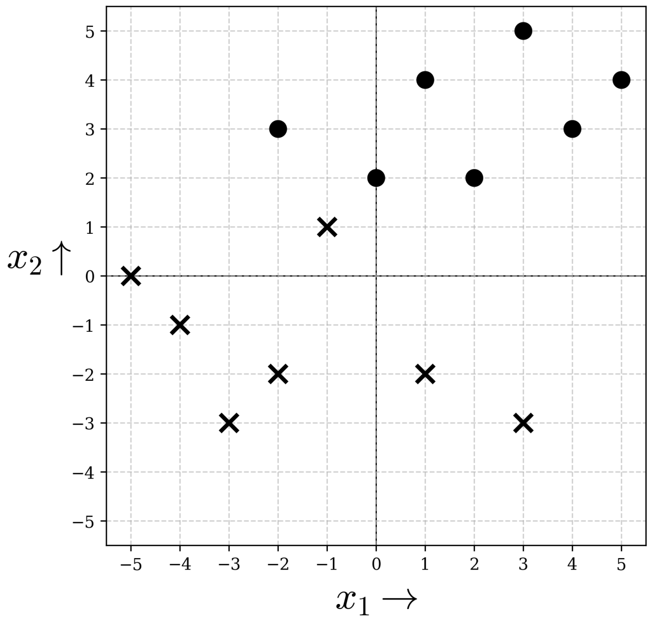
   (b) 3-NN

   (c) 5-NN

   (d) 9-NN

2. 유클리드 거리 대신 맨해튼 거리(Manhattan distance)를 사용하면, LOOCV 오류가 감소하는 모델($k = 1, 3, 5, 9$ 중)이 존재한다. (O / X)

문제 10. **[14점]** 다음의 2차원 데이터에 대한 분류 문제를 로지스틱 회귀 모델로 해결하고자 한다. '$\times$' 표시는 클래스 1 ($y = 1$), '•' 표시는 클래스 0 ($y = 0$)에 해당하며, 전체 학습 데이터셋을 $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{n}$ ($n = 14$) 라고 하자.



사용하는 로지스틱 회귀 모델은 다음과 같다:

$$P(y = 1|\mathbf{x}, \mathbf{w}) = \frac{1}{1 + \exp\left(-(w_0 + w_1 x_1 + w_2 x_2)\right)}$$

위 그림에서 볼 수 있듯, 이 데이터는 선형 분리가 가능하므로 학습 오류(training error)를 0으로 만들 수 있다. 다음 물음에 답하시오.

1. 모델의 과적합을 방지하기 위해, 하나의 가중치 w_j에만 정규화(regularization)를 적용하여 다음과 같은 목적 함수를 최대화하려고 한다: (단, C는 매우 큰 양수이다.)

$$\sum_{(\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}} \log P(y^{(i)}|\mathbf{x}^{(i)}, w_0, w_1, w_2) - C w_j^2 \quad (j \in \{0, 1, 2\})$$

다음의 각 가중치에 정규화를 강하게 적용할 때 (즉, C가 매우 클 때), 정규화가 없을 때와 비교하여 학습 오류의 변화를 "증가", "감소", "유지" 중 하나로 선택하고, 그 이유를 간략히 설명하시오.

(a) 가중치 w_2를 정규화할 경우 ($j = 2$)

(b) 가중치 w_1를 정규화할 경우 ($j = 1$)

(c) 가중치 w_0를 정규화할 경우 ($j = 0$)

2. 이번에는 정규화 방식을 바꾸어, w_1과 w_2에만 정규화를 적용한다고 가정하자. 목적 함수가 다음과 같이 변경되었다:

$$\sum_{(\mathbf{x}^{(i)},y^{(i)})\in\mathcal{D}} \log P(y^{(i)}|\mathbf{x}^{(i)}, w_0, w_1, w_2) - C(|w_1| + |w_2|)$$

정규화 강도 C를 점차 증가시킬 때, 가중치 w_1과 w_2 중 어느 것이 먼저 0이 될 것으로 예상하는가? 또한, 그 이유를 간략히 설명하시오.

문제 11. [**12점**] 어떤 새로운 입자의 에너지 레벨은 모수(parameter) θ를 갖는 레일리 분포 (Rayleigh distribution)를 따른다고 한다. 이 분포의 확률 밀도 함수(Probability Density Function)는 다음과 같다:

$$f(X_i;\theta) = \frac{X_i}{\theta^2}\exp\left(-\frac{X_i^2}{2\theta^2}\right)$$

n개의 표본 $X = (X_1, \ldots, X_n)$을 얻었을 때, 우도 함수(Likelihood Function)는 다음과 같다:

$$P(X \mid \theta) = \prod_{i=1}^{n} f(X_i;\theta)$$

또한, 베이즈 추론을 위해 모수를 $\lambda = \frac{1}{2\theta^2}$로 재정의한다. λ에 대한 우도 함수와 사전 확률은 각각 다음과 같이 비례한다고 알려져 있다: (단, α와 β는 사전 분포의 형태를 결정하는 주어진 양의 상수이다.)

$$P(X \mid \lambda) \propto \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} X_i^2\right), \quad P(\lambda) \propto \lambda^{\alpha-1}\exp(-\beta\lambda)$$

1. 최대우도추정법(MLE)을 사용하여 θ의 추정치 $\hat{\theta}_{MLE}$를 구하시오.
2. 베이즈 정리를 이용하여, 사후 분포 $P(\lambda \mid X)$가 어떤 함수에 비례하는지 구하시오.
3. 최대사후확률(MAP) 추정법을 사용하여 λ의 추정치 $\hat{\lambda}_{MAP}$를 구하시오.