기계학습원론 기말고사
2024 가을

강의자: 조재민
강의조교: 이재웅, 정명원

December 14, 2024

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

1. 비볼록 함수(non-convex fuction)에서 어떤 지점의 그래디언트가 0이면, 그 지점은 전역 최솟값(global minimum) 또는 지역 최솟값(local minimum)이다.
2. 입력 벡터의 모든 성분을 동일한 양의 상수로 나눈 뒤, 동일한 상수를 빼더라도 소프트맥스 함수를 적용한 결괏값은 변하지 않는다.
3. 활성화 함수 $f(x) = ax + b$ $(a, b, x \in \mathbb{R})$를 사용하는 다층 퍼셉트론에서, 충분한 수의 은닉 뉴런과 깊은 층을 가진다면 임의의 비선형 함수를 근사할 수 있다.
4. F1-score는 Precision과 Recall의 산술평균보다 항상 크거나 같다.
5. Bagging 기법은 편향은 크게 줄이지 못하지만 분산은 감소시키는 반면, Boosting 기법은 편향과 분산을 모두 줄이는 경향이 있다.
6. 계층적 군집화(hierarchical clustering)는 연결 기준(linkage criteria)을 통해 군집 간 거리를 정의하고, DBSCAN은 밀도 기반으로 군집을 형성하며, 두 방법 모두 군집 수를 사전에 지정할 필요가 없다.
7. PCA와 같은 선형 차원 축소 기법은 주어진 데이터에 대해 선형 변환을 통해 저차원 표현을 얻는 반면, t-SNE와 같은 비선형 차원 축소 기법은 데이터가 내재한 비선형적 매니폴드 구조를 유지하도록 임베딩하여 저차원 표현을 찾는다.
8. 로지스틱 회귀의 비용 함수는 파라미터에 대해 볼록(convex)하지만, 닫힌 형태의 해(closed-form solution)는 존재하지 않는다.
9. 가우시안 혼합 모델(Gaussian Mixture Model) 분류기는 각 데이터 포인트를 소프트 할당(soft assignment) 방식으로 처리하는 생성적 분류기(generative classifier)이다.
10. 다음은 Soft-Margin SVM의 목적함수이다:

$$\min_{w,b,\{\xi_i\}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \quad \text{subject to} \quad \begin{cases} y^{(i)}(w^\top x^{(i)} + b) \geq 1 - \xi_i \quad (\forall i), \\ \xi_i \geq 0 \quad (\forall i) \end{cases}$$

라그랑지안을 정의한 뒤 최적화하여 얻은 쌍대 문제(dual problem)는 다음과 같다:

$$\mathcal{L}(w, b, \{\xi_i\}, \{\alpha_i\}, \{\mu_i\}) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\big(y^{(i)}(w^\top x^{(i)} + b) - 1 + \xi_i\big) - \sum_{i=1}^{n}\mu_i\xi_i$$

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y^{(i)}y^{(j)}(x^{(i)})^\top x^{(j)} \quad \text{subject to} \quad \begin{cases} 0 \leq \alpha_i \leq C \quad (\forall i), \\ \sum_{i=1}^{n}\alpha_i y^{(i)} = 0 \end{cases}$$

이때, $0 < \alpha_i < C$인 데이터 포인트는 반드시 마진을 결정하는 두 초평면 중 하나 위에 위치한다. (단, 모든 수식 도출은 올바르다고 보아도 좋다.)

문제 2. 길이가 2인 2차원 벡터 x와 양의 정부호(positive-definite) 2×2 행렬 A에 대해 $f(x) = x^T Ax$를 최소화 하는 x의 조건과 그 때의 최솟값을 A를 활용하여 설명하여라. 완결된 식이 아니더라도 좋다. 그리고 이 결과를 이용하여 $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$에 대해 $f(x)$의 최솟값을 구하여라.

문제 3. 다음과 같은 데이터가 주어져 있다.

$$(x_1, y_1) = (1, 2), \quad (x_2, y_2) = (2, 5), \quad (x_3, y_3) = (3, 7)$$

1. 오차를 실제 값과 예측 값의 제곱으로 두고, 오차의 합을 최소화하는 선형 회귀 직선 $y = w_0 + w_1 x$을 편미분을 사용하여 구하여라.
2. 정규 방정식(normal equation)을 사용하여 선형 회귀 직선 $y = w_0 + w_1 x$을 구하여라.

문제 4. 다음 표는 날씨, 시간대에 따른 교통사고 발생 여부를 나타내는 데이터셋이다. 계산 시 모든 로그(log) 연산은 밑을 2로 사용하라.

날씨	시간대	교통사고
맑음	주간	False
맑음	주간	False
흐림	주간	True
흐림	야간	False
비	야간	True
비	야간	False
눈	주간	True
눈	야간	True

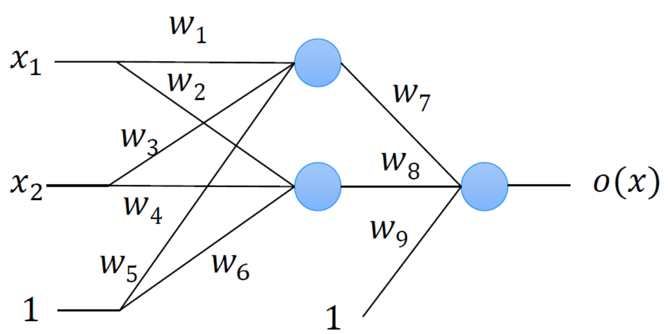
1. 날짜와 시간대 두 특성 각각에 대해 Information Gain(*InfoGain*)을 구하시오.
2. 두 특성 각각에 대해 Split Information(*SplitInfo*)을 구한 뒤, 이를 활용하여 각 특성의 *GainInfo*를 계산하시오.

문제 5. 다음 데이터셋에 AdaBoost 모델을 학습하려한다. 약한 분류기(weak learner)는 데이터의 실제 레이블(label)인 y에 대해, 첫 번째 라운드와 두 번째 라운드에서 샘플을 각각 아래 표의 y_1과 y_2와 같이 예측한다. 아래의 각 샘플의 초기 가중치를 보고 이후 각 라운드에서 갱신되는 가중치를 계산하시오. (단, 갱신된 가중치는 정규화되어야 한다.)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$	$-$	$-$
y_i	$+$	$+$	$+$	$-$	$+$	$-$	$-$	$-$	$-$	$-$
y_2	$+$	$+$	$+$	$-$	$-$	$-$	$-$	$-$	$-$	$-$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
초기 가중치	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10
라운드 1 이후 가중치										
라운드 2 이후 가중치										

문제 6. 다음과 같이 2개의 은닉 뉴런 h_1, h_2를 가지는 다층 신경망을 고려하자:



$$f(z) = \begin{cases} 1 & \text{if } z > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$h_1 = f(x_1 w_1 + x_2 w_3 + w_5)$$

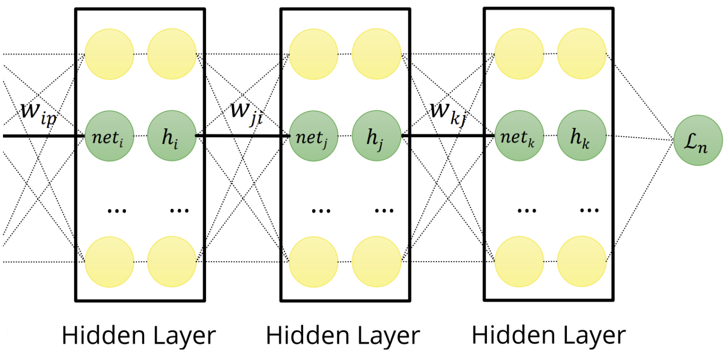
$$h_2 = f(x_1 w_2 + x_2 w_4 + w_6)$$

$$o = f(h_1 w_7 + h_2 w_8 + w_9)$$

이 때, $w_1 = w_2 = w_3 = w_4 = 1.0, w_5 = -0.5, w_6 = -1.5, w_7 = 1.0, w_8 = -1.0$ 이다.

1. $x_1 = 1, x_2 = 1$일 때, h_1의 값을 구하시오.
2. $x_1 = 0, x_2 = 0$일 때, h_2의 값을 구하시오.
3. $(x_1, x_2) \in \{0, 1\}^2$에 대하여 $o(x_1, x_2) = x_1 \oplus x_2$를 만족하기 위한 w_9의 범위를 구하시오. (단, $\oplus$는 XOR 연산이다.)

문제 7. 다음과 같은 다층 신경망 구조를 고려하자:



$$\mathcal{L}_n = \frac{1}{2} \sum_k (t_k - h_k)^2 \quad (t_k : \text{target})$$

$$\text{net}_k = \sum_j h_j w_{kj}, \quad h_k = \sigma(\text{net}_k)$$

$$\text{net}_j = \sum_i h_i w_{ji}, \quad h_j = \sigma(\text{net}_j)$$

$$\text{net}_i = \sum_p h_p w_{ip}, \quad h_i = \sigma(\text{net}_i)$$

이 때, $\sigma(\cdot)$은 시그모이드 함수이며, 연쇄 법칙에 따라 다음이 성립한다:

$$\frac{\partial \mathcal{L}_n}{\partial w_{kj}} = \frac{\partial \mathcal{L}_n}{\partial \text{net}_k} \frac{\partial \text{net}_k}{\partial w_{kj}} = \delta_k h_j$$

$$\frac{\partial \mathcal{L}_n}{\partial w_{ji}} = \frac{\partial \mathcal{L}_n}{\partial \text{net}_j} \frac{\partial \text{net}_j}{\partial w_{ji}} = \delta_j h_i$$

$$\frac{\partial \mathcal{L}_n}{\partial w_{ip}} = \frac{\partial \mathcal{L}_n}{\partial \text{net}_i} \frac{\partial \text{net}_i}{\partial w_{ip}} = \delta_i h_p$$

1. δ_k를 t_k와 h_k에 관한 식으로 표현하시오.
2. δ_j를 δ_k, w_{kj}, h_j 에 관한 식으로 표현하시오.
3. δ_i를 δ_j, w_{ji}, h_i 에 관한 식으로 표현하시오.

문제 8. 다음과 같은 2차원 데이터 포인트들이 주어져 있다.

$$x_1 = (6, 8), \quad x_2 = (1, 3), \quad x_3 = (3, 2), \quad x_4 = (2, 1), \quad x_5 = (7, 6), \quad x_6 = (8, 7).$$

k-Means 알고리즘을 이용하여 $k = 2$개의 군집 C_1과 C_2로 데이터를 분할하기 위해 초기 군집 중심(centroid)을 각각 $c_1^{(0)} = (3, 3), \quad c_2^{(0)} = (6, 6)$와 같이 설정했다.

1. 첫번째 할당 단계와 업데이트 단계를 마친 뒤 계산된 새로운 군집 중심 좌표 $c_1^{(1)}, c_2^{(1)}$를 구하시오.
2. 첫번째 할당 단계 이후 누락된 데이터 $x_7 = (3, 5), \quad x_8 = (6, 4)$를 추가하여 할당과 업데이트 단계를 반복했을 때 계산되는 군집 중심 좌표 $c_1^{(2)}, c_2^{(2)}$를 구하시오.

문제 9. 다음 데이터를 이진 분류하려고 한다:

$$+1 \text{ 클래스: } (1, 1), (2, 2) \qquad \text{-1 클래스: } (1, -1), (2, -2)$$

이 데이터를 분류하기 위해 다음 최적화 문제를 고려하자:

$$\min_{\mathbf{w}, b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad (i = 1, 2, 3, 4)$$

이 때, $\mathbf{x} = [x_1, x_2]^\top$는 입력 벡터, $y_i \in \{+1, -1\}$는 각 데이터 점 $\mathbf{x}_i$의 클래스 레이블, $\mathbf{w}$는 분류기 파라미터(가중치 벡터), b는 편향 파라미터이다.

1. 위의 최적화 문제를 만족하는 결정 경계를 x_1, x_2에 관한 식으로 표현하시오.
2. 사상 함수(mapping function)를 $\phi(\mathbf{x}) = [1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2]^\top$로 정의하자. $\mathbf{x}' = [x_1', x_2']^\top$에 관한 커널 함수 $k(\mathbf{x}, \mathbf{x}')$를 $\mathbf{x}$와 $\mathbf{x}'$에 관한 식으로 표현하시오.