

14. (1 점: 문항당 0.25 점)

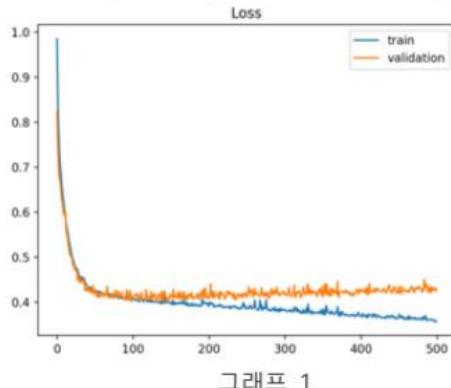
Overfitting 을 방지하기 위해 weight penalty(가중치 벌칙)로 L1-norm 을 사용하는 경우와 L2-norm 을 사용하는 경우의 공통점과 차이점에 대해 설명한 것이다.

각 설명에 대해 맞으면 O, 틀리면 X로 표시 하시오.

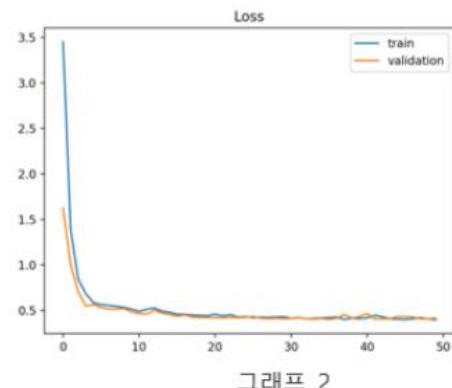
- (1) L1 과 L2 모두 가중치의 크기를 0에 가깝게 하는 효과가 있다.
- (2) L1 과 L2 모두 bias 에 해당하는 가중치에는 weight penalty 를 적용하지 말아야 한다.
- (3) 가중치 업데이트시 weight penalty 에 의해 변화되는 가중치의 크기는, L1 의 경우 현재 가중치 크기와 무관하게 일정하지만 L2 의 경우는 현재 가중치 크기에 비례한다.
- (4) L1 은 L2 보다 non-zero 가중치가 더 많이 발생하는 경향이 있다.

15. (2 점: 빈칸당 0.5 점)

다음 그라프들은 training set 과 validation set 에 대한 기계학습 모델의 학습곡선(learning curve)들을 나타낸다. (학습곡선에서 가로축은 epoch 수를, 세로축은 매 epoch 수행수의 loss 값을 나타낸다.) 각 그라프는 서로 다른 데이터셋과 모델을 사용한 결과이며, 각 그라프에 나타낸 최종 epoch 만큼 해당 모델을 학습 시켰다고 한다. 그라프 1~2 각각의 최종학습상태와 개선방안을 각각 <보기 1>과 <보기 2>에서 고르되, 해당되는 기호를 모두 고르시오.



그래프 1



그래프 2

<보기 1. 학습된 모델의 상태>

- A. 이 모델은 과소적합(underfit)일 가능성이 높다.
- B. 이 모델은 과대적합(overfit)일 가능성이 높다.
- C. 이 모델은 적합(good fit) 상태일 가능성이 높다.
- D. 이 모델은 high-bias 일 가능성이 높다.
- E. 이 모델은 high-variance 일 가능성이 높다.

<보기 2. 개선 방안>

- 가. Epoch 수를 늘려 학습을 더 시키는 게 도움이 된다.
- 나. Training data 를 늘려 다시 학습 시키는 게 도움이 된다.
- 다. Weight penalty 와 같은 regularization(규제) 방법을 사용하여 다시 학습 시키는 게 도움이 된다.
- 라. 모델의 용량(capacity)를 늘려 다시 학습 시키는 게 도움이 된다.
- 마. 현재 상태로 학습을 완료하는 게 좋다.
- 바. Validation loss 가 가장 작은 epoch 에서 학습을 조기종료(Early Stop)하는 게 도움이 된다.

16.

(1) 3-way hold-out 방법을 사용한 모델 선택 및 평가 과정에 대해 설명하시오.

(2) 5-fold cross-validation 방법을 사용한 모델 선택 과정에 대해 설명하시오.

(3) 3-way hold-out 방법과 k-fold cross-validation 방법은 각각 어떤 상황에서 적용하면 좋을지 설명하시오.

문제1. (1.5점: 빙칸당 0.5) SVM을 이용한 분류에 관한 다음 설명에 대해 옳으면 O, 틀리면 X로 표시하시오.

1) 선형분리 가능한 훈련데이터에 대해 hard-margin SVM의 결정경계와 soft-margin SVM의 결정경계는 항상 동일하다.

2) One-versus-Rest 방식으로 multi-class classification 문제를 풀고자 한다면, class 개수와 동일한 개수의 이진 SVM이 필요하다.

3) SVM의 결정 경계 $d(x)$ 에 서포트 벡터를 대입하면 항상 0이된다.

문제2. (1점) 커널 함수란 무엇인지 정의하시오.

문제3. (1점: 문항당 0.5) 커널 트릭이란 무엇인지 정의하고, 공간변환과 비교하여 장점을 기재하시오.

1) 커널 트릭:

2) 공간변환 대비 장점:

문제4. (2점: 문항당 0.5) 군집의 개수 k 를 알고 있다고 가정하자. 이때 k-Means 알고리즘을 이용한 군집화의 한계와 개선 방안(K-medoids 제외)에 대해 설명하시오.

1) 한계:

2) 개선 방안:

3) 군집의 개수 k 를 모르는 경우를 가정하자. 이때 k 를 결정하기 위해 Elbow Method를 사용하고자 한다. 그래프를 그려 elbow 지점을 찾고자 할 때, 이 그래프의 가로축과 세로축은 각각 무엇이 되어야 하는가? (둘 다 맞아야 0.5점)

4) k-means와 k-medoids의 차이점에 대해 설명하시오.

문제5. (3점: 문항당 1) 다음 각 비모수적 확률밀도 추정 방법에 대해 설명시오.

1. k-Nearest Neighbors (k-NN)
2. Parzen Window
3. Kernel Density Estimation (KDE)

문제6. [풀이과정 필요] (5점: 빙칸당 0.5)

훈련집합 $\mathbb{X} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), (\mathbf{x}_3, y_3), (\mathbf{x}_4, y_4)\}$ 가 다음과 같이 주어졌다.

Hard-Margin Linear SVM을 이용하여 이진 분류기를 설계하고자 할 때,
물음에 답하시오. (아래에서, $\mathbf{x} = (x_1 \ x_2)^T$)

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad y_1 = y_2 = 1, \quad y_3 = y_4 = -1$$

1.

x_1 축과 이루는 각도가 0도, 45도, -45도, 90도인 직선들 중에서 최적 결정 경계를 찾고자 한다. 찾아낸 최적 결정 경계 방정식이 다음 식과 같을 때, 상수 A~C의 값을 적으시오. ($d(\mathbf{x})$ 가 양수이면 -1로, 음수이면 1로 분류한다고 가정하시오.)

(Hint: 특징공간에서 그림을 그리면 최적 결정 경계를 쉽게 찾을 수 있음)

$$d(\mathbf{x}) = Ax_1 + Bx_2 + C = 0$$

$$A = \boxed{}, \quad B = \boxed{}, \quad C = \boxed{}$$

2. 위에서 찾은 결정 경계로부터 SVM의 여백(margin)의 크기를 구하시오.

$$\text{margin} = \boxed{}$$

3. 이번에는 주어진 SVM 문제를 아래와 같은 Wolfe dual 문제로 바꿔서 풀고자 한다.

$$\text{Maximize: } \tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{Subject to: } \sum_{i=1}^n \alpha_i y_i = 0 \quad 0 \leq \alpha_i, \quad i = 1, 2, \dots, n.$$

Support vector가 \mathbf{x}_1 과 \mathbf{x}_4 라는 조건하에서 Wolfe dual 문제를 풀어 최적의 $\alpha = (\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4)^T$ 를 구한 후, α_1 과 α_2 의 값을 적으시오.

$$\alpha_1 = \boxed{}, \quad \alpha_2 = \boxed{}$$

위 3의 결과를 이용하여 구한 결정 경계의 방정식이 다음 식과 같을 때,

4. 상수 D~F의 값을 적으시오.

$$d(\mathbf{x}) = Dx_1 + Ex_2 + F = 0$$

$$D = \boxed{}, \quad E = \boxed{}, \quad F = \boxed{}$$

5. 위 3~4를 통해 설계된 SVM에 새로운 샘플 $\mathbf{x} = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}$ 를 입력하였을 때,
SVM의 출력 값을 구하시오.

$$\text{출력값} = \boxed{}$$

문제7. [풀이과정 필요] (2점: 빈칸당 1)

앞의 문제6에서의 훈련집합 \mathcal{X} 에 다음과 같은 두 개의 샘플이 추가되었다고 한다.
Soft-Margin SVM을 이용하여 이진 분류기를 설계하고자 할 때, 물음에 답하시오.

$$\mathbf{x}_5 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad y_5 = 1, \quad y_6 = -1$$

Soft-Margin SVM의 비용함수 $J(\mathbf{w}, \xi)$ 는 아래 식과 같다.

$C = 1$ 이라 할 때, 아래 주어진 결정 경계 d_1 과 d_2 각각에 대한

최저 비용함수 값을 소수 첫째자리까지 계산하시오.

(Hint: 특징공간에서 그림을 그리면 쉽게 계산할 수 있음)

$$J(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \xi_i \quad \begin{aligned} d_1(\mathbf{x}) &= -x_2 + 1 = 0 \\ d_2(\mathbf{x}) &= x_1 - x_2 = 0 \end{aligned}$$

문제8. [풀이과정 필요] (4점: 빙칸당 1)

어떤 동네에 애완동물이 총 12마리 있다고 한다.

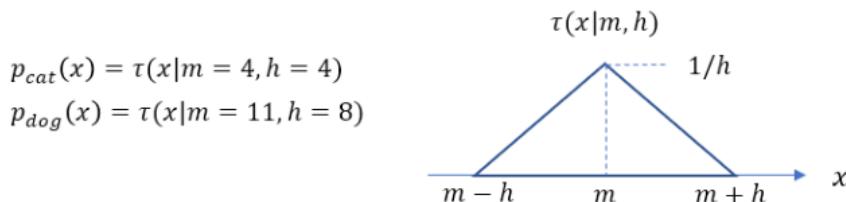
12마리로부터 털을 각각 하나씩 채취하여 길이를 측정하였더니 아래와 같았다.

아래에서 x_i 는 i번째 애완동물의 털 샘플 길이(cm)를 나타내며,
 $x_1 \sim x_8$ 은 cat으로부터, $x_9 \sim x_{12}$ 는 dog으로부터 채취됐다고 한다.

$$x_1 = x_2 = 2, \quad x_3 = x_4 = x_5 = x_6 = 4, \quad x_7 = x_8 = 6$$

$$x_9 = 7, \quad x_{10} = x_{11} = 11, \quad x_{12} = 15$$

Cat과 dog의 털길이 확률분포를 각각 $p_{cat}(x)$ 와 $p_{dog}(x)$ 라 하고,
 $p_{cat}(x)$ 와 $p_{dog}(x)$ 이 각각 아래와 같은 모수적 확률분포 $\tau(x|m, h)$ 를 따른다고 한다.



길을 걷다가 주운 털의 길이가 7cm (즉, $x = 7$)였다고 할 때,
이 털이 cat의 털인지 dog의 털인지를 판별하고자 한다. 다음 물음에 답하시오.

1. ML (Maximum Likelihood) 방법으로 판별하기 위해 likelihood를 구하고자 한다.

주운 털 x 에 대한 cat과 dog의 likelihood를 각각 구하시오(기약분수로 답하시오).

cat의 likelihood = , dog의 likelihood =

2. cat의 사전확률 $p(cat)$ 을 주어진 12개의 샘플들로부터 추정하시오(기약분수로 답하시오).

$p(cat) =$

3. MAP (Maximum A Posteriori) 방법으로 판별한 결과를 아래에서 골라 기호로 답하시오: 기호 =

- A. 주운 털은 cat의 것이다.
- B. 주운 털은 dog의 것이다.
- C. 주운 털이 cat의 것인지 dog의 것인지 사후확률이 같아 판별할 수 없다.

문제9. [풀이과정 필요] (5점: 빈칸당 1)

문제8에서와 동일하게 주어진 12개의 털 샘플들을 길이에 따라 두개의 군집 A와 B로 나누고자 한다.
(단, 주어진 털이 cat의 것인지 dog의 것인지 모른다고 가정하라)

군집화를 위해 **K-Means** 알고리즘을 이용하고자 하며, 군집 A와 B의 초기 군집중심이 각각 $z_A=3$, $z_B=8$ 이라 하자.

1. 초기 군집중심에 의해 샘플들을 군집 A와 B에 배정하였다.

이때 군집 A에 배정된 샘플은 몇 개인가? 개

2. 위 1의 군집화 결과를 이용하여 군집중심을 새로 구하였다.

새 군집중심 z_A 와 z_B 는 각각 얼마인가? (소수 첫째 자리까지만 구하시오)

$z_A =$, $z_B =$

3. 위 2에서 구한 군집 중심을 이용하여 샘플들을 군집 A와 B에 다시 배정한 후, 배정 결과를 이용하여 군집 중심을 또 다시 구하였다.

또 다시 구한 군집중심 z_A 와 z_B 는 각각 얼마인가? (소수 첫째 자리까지만 구하시오)

$z_A =$, $z_B =$

문제10. [풀이과정 필요] (6점: 빙칸당 1)

문제8에서와 동일하게 주어진 12개의 텔 샘플들의 확률분포를 2개의 가우시언 분포로 구성된 GMM 으로로 추정하고자 한다.

EM(Expectation Maximization) 알고리즘 적용을 위해, 첫번째 분포와 두번째 분포의 파라메터들을 난수로 초기화 하였더니, 다음과 같았다.

- 각 분포의 평균: $\mu_1=3, \mu_2=8$
- 각 분포의 분산: $\sigma_1^2=4, \sigma_2^2=4$
- 각 분포의 발생확률: $\pi_1=4/5, \pi_2=1/5$

- 소속정보 행렬 $\mathbf{Z} = [z_{ki}]$ 의 원소 z_{ki} 를 샘플 \mathbf{x}_i 가 k^{th} 분포에 속할 확률로 정의할 때, 위의 초기 파라메터들로부터 $z_{1,3}$ 와 $z_{2,8}$ 의 값을 각각 구하시오(기약분수로 답하시오).

$$z_{1,3} = \boxed{}, \quad z_{2,8} = \boxed{}$$

- EM 알고리즘을 여러번 반복 수행 중에 프로그램을 중지하였더니 소속정보 행렬 \mathbf{Z} 가 다음과 같았다고 한다.
 \mathbf{Z} 를 이용하여 두 가우시언 분포의 평균과 분산을 ML 방식으로 추정하시오.

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

- 각 분포의 평균: $\mu_1 = \boxed{}, \mu_2 = \boxed{}$
- 각 분포의 분산: $\sigma_1^2 = \boxed{}, \sigma_2^2 = \boxed{}$

[Hint: 공식이 기억이 안나면 아래 수식을 참고해도 됩니다.]

$$z_{ki} = \frac{\pi_k N(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j N(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

$$\boldsymbol{\mu}_k = \frac{1}{n_k} \sum_{i=1}^n z_{ki} \mathbf{x}_i$$

$$\boldsymbol{\Sigma}_k = \frac{1}{n_k} \sum_{i=1}^n z_{ki} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$$

$$\pi_k = \frac{n_k}{n}$$

$$\text{where } n_k = \sum_{i=1}^n z_{ki}$$

문제11. [풀이과정 필요] (6점: 빙칸당 0.5)

다음은 어떤 식물의 잎의 개수와 꽃의 개수를 8개월에 걸쳐 관측한 데이터이다.

$$\mathbf{x}_1 = \begin{pmatrix} 10 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \mathbf{x}_7 = \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \mathbf{x}_8 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$$

여기서 \mathbf{x}_i 는 i 번째 달에 관측한 잎의 개수와 꽃의 개수로 이루어진 2차원 특징벡터를 의미한다.
(즉, $\mathbf{x} = [x_1 \ x_2]^T$, x_1 : 잎의 개수, x_2 : 꽃의 개수)

주성분분석(PCA)을 이용하여 주어진 특징공간의 차원을 2차원에서 1차원으로 축소하고자 할 때,
다음 물음에 답하시오. (주의: Centering 및 그 역과정을 적절히 사용해야만 올바른 결과를 얻을 수 있음)

- 1) 훈련집합 $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_8\}$ 로부터 특징벡터 \mathbf{x} 의 평균 $\boldsymbol{\mu} = [\mu_1 \ \mu_2]^T$ 를 추정하시오.

$$\mu_1 =$$

$$\mu_2 =$$

- 2) PCA 적용을 위해 필요한 공분산행렬 K 를 추정한 후, 첫번째 행의 원소(즉, σ_{11} 및 σ_{12})의 값을 기재하시오.

$$\sigma_{11} =$$

$$\sigma_{12} =$$

- 3) 잎의 개수와 꽃의 개수 간의 상관계수(correlation coefficient)를 계산하시오(기약분수로 답하시오).

$$\rho =$$

- 4) 2)에서 구한 공분산행렬 K 의 고유값(eigenvalue)들을 구해 크기 순서대로 기재하시오.

$$\lambda_1 =$$

$$\lambda_2 =$$

- 5) 4)에서 구한 고유값 λ_1 에 해당하는 고유벡터를 구했더니 다음과 같았다고 한다.
기호 a, b 에 알맞은 값을 기재하시오.

$$\mathbf{u}_1 = \frac{1}{\sqrt{a}} \begin{pmatrix} 1 \\ b \end{pmatrix}$$

$$a =$$

$$b =$$

- 6) 위의 결과들을 이용하면 특징공간을 2차원에서 1차원으로 변환하는 PCA 변환(encoding) 행렬과
역변환(decoding) 행렬을 구할 수 있다.

새로운 관측치 $\mathbf{x} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$ 가 주어졌을 때 대한 PCA 변환(encoding) 결과 z 를 구하시오:

$$z =$$

- 7) 6)에서 구한 z 를 PCA 역변환(decoding)하여 최종 복원한 결과를 \mathbf{x}' 를 구했더니 다음과 같았다고 한다.
기호 c, d 에 알맞은 값을 기재하시오. $\mathbf{x}' = \begin{pmatrix} c \\ d \end{pmatrix}$

$$c =$$

$$d =$$