

인공지능 (202220HY23385) 기말고사 (2022년 2학기)

담당교수: 정우환

1. 다음 O/X 문제에 대답하시오. (각 2점)

- (1) SVM (Support Vector Machine)에 Soft-margin을 적용하면 XOR문제와 같은 Linearly non-separable 문제를 풀 수 있다.

Answer) X

- (2) K-NN classifier는 학습시간이 오래 걸린다는 단점이 있다.

Answer) X

- (3) Validation data는 최적의 learning rate 결정에 활용할 수 있다.

Answer) O

- (4) Test data는 최적의 hidden layer 개수를 결정할 때 활용할 수 있다.

Answer) X

- (5) Hidden layer의 activation function으로 Tanh 대신 ReLU를 사용하면 Vanishing gradient 문제를 줄일 수 있다.

Answer) O

2. 1과 0중 항상 1 (Positive) 로 대답하는 Binary classification 알고리즘이 있다. 테스트셋에 있는 100개의 데이터 중 90개의 정답이 1일때, 아래 질문에 답하여라. (각 3점)

- (1) Precision을 계산하라.

Answer) 0.9

- (2) Recall을 계산하라.

Answer) 1.0

- (3) F1-score를 계산하라. (기약분수로 표현하시오)

Answer) $2 \frac{1.0 \times 0.9}{1.0 + 0.9} = \frac{18}{19}$

- (4) Precision, Recall, F1-score가 해당 알고리즘의 정확도를 측정하는데 충분한지 이야기하고 부족하다면

추가로 사용할 수 있는 Measure를 이야기하시오.

Answer) 항상 1이라고 대답하는 ZeroR알고리즘에 지나치게 높은 score를 부여하기 때문에 부족함.

$$\text{Specificity} = (\text{TN})/(\text{TN}+\text{FP}) = 0$$

3. 어떤 집의 특성을 나타내는 입력데이터 $\mathbf{x} \in \mathbb{R}^d$ 가 주어졌을 때, 그 집의 가격 y 를 추정하기 위한 회귀 모델을 학습시키고자 한다. 집값은 양수라는 사실을 반영하기 위해 선형회귀 (Linear regression) 모델을 출력값을 ReLU 함수에 통과시켜 항상 양의 실수를 출력하는 모델을 다음과 같이 구성하였다.

$$\hat{y} = \text{ReLU}(\mathbf{w}^\top \mathbf{x} + b)$$

Loss함수로 $(y - \hat{y})^2$ 를 사용할 때 아래 물음에 답하여라.

- (1) ReLU함수의 정의를 쓰시오. (5점)

$$\text{Answer) } \text{ReLU}(x) = \text{Max}(0, x)$$

- (2) 다음 편미분 값을 구하여라 $\frac{\partial L(\hat{y}, y)}{\partial w_i}$. (10점)

$$\text{Answer) Let } z = \mathbf{w}^\top \mathbf{x} + b$$

$$\frac{\partial L(\hat{y}, y)}{\partial w_i} = \frac{\partial L(\hat{y}, y)}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial w_i} = \begin{cases} 2(\hat{y} - y)x_i & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{cases}$$

4. Neural network를 이용해 n-ary classification모델을 학습시킬 때는 주로 Cross entropy loss를 많이 이용한다. n 개의 class가 있을 때, 정답 레이블 $\mathbf{y} = [y_1, y_2, \dots, y_n]$ 와 추정값 $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]$ 사이의 cross entropy loss $L(\mathbf{y}, \hat{\mathbf{y}})$ 을 적어라. (5점)

Answer)

$$L(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^n y_i \log \hat{y}_i$$

5. 아래 PyTorch 코드에서 에러를 찾고 수정하여라. (5점)

```
class MLP_h(nn.Module):
    def __init__(self, hidden_units = [512, 256, 128]):
        super().__init__()

        self.in_dim = 28*28 # MNIST
        self.out_dim = 10

        self.l_layers = []
        self.l_layers.append(nn.Linear(self.in_dim, hidden_units[0]))
        for i in range(len(hidden_units)-1):
            self.l_layers.append(nn.Linear(hidden_units[i], hidden_units[i+1]))
        self.l_layers.append(nn.Linear(hidden_units[-1], self.out_dim))

        self.relu = nn.ReLU()
        self.log_softmax = nn.LogSoftmax()

    def forward(self, x):
        a = x.view(-1, self.in_dim)
        for l in range(len(self.l_layers)):
            z = self.l_layers[l](a)
            if l == len(self.l_layers) - 1:
                logit = z
            else:
                a = self.relu(z)

        return logit
```

```
model = MLP_h([2, 3])
criterion = nn.CrossEntropyLoss()
optimizer = optim.SGD(model.parameters(), lr = 0.01)
```

Answer)

Self.l_layers = nn.ModuleList(self.l_layers)

6. d차원 벡터 $\mathbf{x} = (x_1, x_2, \dots, x_d)$ 를 입력으로 받아 n-ary classification을 하는 Naïve Bayes Classifier가 있다 ($y \in \{C_1, C_2, \dots, C_n\}$). 다음 물음에 답하여라

- (1) Naïve bayes classifier는 Posterior probability $P(C_i|\mathbf{X})$ 를 최대화 하는 C_i 를 찾아내는것을 목표로 한다. 이때 $P(C_i|\mathbf{X})$ 를 구하기 위해 아래의 비례식을 이용한다

$$P(C_i|\mathbf{X}) \propto P(\mathbf{X}|C_i)P(C_i)$$

위의 식을 유도 하시오. (5점)

Answer)

By Bayes' theorem에 의해

$$P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i) \frac{P(C_i)}{P(\mathbf{X})}$$

$P(\mathbf{X})$ 는 모든 클래스에 대해 constant이기 때문에,

$$P(C_i|\mathbf{X}) \propto P(\mathbf{X}|C_i)P(C_i)$$

Outlook	Temp	Humidity	Wind	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

- (2) 위의 식에서 $P(\mathbf{X}|C_i)$ 를 계산하기 위한 식을 쓰시오. (5점)

Answer)

Conditional independence 가정에 의해 $P(\mathbf{X}|C_i) = \prod_{j=1}^d P(X_j|C_i)$

- (3) 오른쪽 데이터를 활용하여 $P(C_{Yes})$ 를 계산하시오. (5점)

$$P(C_{Yes}) = \frac{9}{14}$$

Figure 1 Play tennis 데이터

7. 다음과 같은 데이터셋으로 Decision tree classifier를 만들고자 한다. 다음 물음에 답하시오. (단, $\log_2 3 \approx 1.58$)

Price	Maintenance	Safety	Label
low	high	Low	negative
high	med	Low	negative
low	low	High	positive
low	high	high	positive

- (1) 위 데이터의 Information $I(D)$ 를 구하시오. (5점)

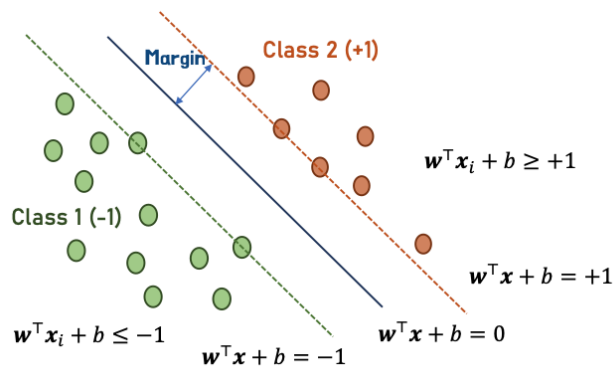
$$Info(D) = I(2,2) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

- (2) Safety attribute로 데이터를 나눴을때의 Information $I_{Safety}(D)$ 를 구하시오. (5점)

$$Info_{safety}(D) = \frac{2}{4}I(2,0) + \frac{2}{4}I(0,2) = 0$$

8. Support vector machine(이하 SVM)은 서로 다른 두 class에 속한 dataset을 가장 큰 margin으로 분리하는 hyperplane을 찾는 문제이고 아래와 같이 정의할 수 있다.

Data : $\langle x_i, y_i \rangle$ for $i = 1, \dots, n$ where $x_i \in \mathbb{R}^d, y_i \in \{-1, +1\}$
Goal : Finding a good separating hyperplane, $w^T x + b = 0$



다음 물음에 답하시오.

- (1) 아래는 SVM의 문제를 constrained optimization problem으로 정의한 수식이다. $[A], [B]$ 에 알맞은 수식을 구하시오. (5점)

$$\min_{w,b} [A]$$

Subject to [B], $i = 1, 2, \dots, n$

Answer)

$$[A]: \frac{1}{2} \|w\|_2^2$$

$$[B]: y_i(w^\top x_i + b) \geq 1$$

(2) 문제 (2)의 Optimization problem에 Soft margin을 추가한 버전에 대한 문제정의를 작성하시오. (5점)

Answer)

$$\min_{w,b} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i$$

Subject to

$$y_i(w^\top x_i + b) \geq 1 - \sum_{i=1}^n \xi_i, \text{ for } i = 1, 2, \dots, n$$

$$\xi_i \geq 0, \text{ for } i = 1, 2, \dots, n$$

(3) 1차원 입력 데이터 x 를 3차원으로 mapping하는 함수 $\Phi(x)$ 가 있다.

$$\Phi(x) = (x^2, \sqrt{2}x, 1)$$

이에 해당되는 Kernel함수를 $K(x_i, x_j)$ 를 쓰시오. . (5점)

Answer)

$$K(x_i, x_j) = \Phi(x_i)^\top \Phi(x_j) = x_i^2 x_j^2 + 2x_i x_j + 1$$