|  |  |
| --- | --- |
| [Machine Learning]  [2021-1] |  |
| Homework 1 |  |
| [Due Date] 2021.04.02  Student ID : 2016112158  Name : KimHeeSu  Professor : Juntae Kim | logo-placeholder |

1. Write python codes to solve each of the following problem, and attach the result and description. (20 pts)

* 1. Python : Circle and Rectangle Class design (Week02-Quiz4)

|  |
| --- |
| Code |
| class Circle:  def \_\_init\_\_(self, radius):  self.radius = radius  def area(self):  return self.radius \*\* 2 \* 3.14  # make Rectangle class  class Rectangle:  def \_\_init\_\_(self, width, height):  self.width = width  self.height = height  def area(self):  return self.width \* self.height |
| Result(Captured images) |
|  |
| Description |
| Circle 클래스의 생성자는 radius를 파라미터로 받고, 그 값으로 self.radius(멤버)를 초기화한다. Circle 클래스의 area메소드는 self.radius로 원의 넓이를 계산하는 메소드이다.  Rectangle 클래스의 생성자는 height(높이)와 (width)를 파라미터로 받고 그 값으로 self.height와 self.width를 초기화한다. area메소드는 self.width와 self.height로 직사각형의 넓이를 계산하는 메소드이다. |

1-2. Numpy : Matrix Dot Product

For , ,

Compute where

Use these functions:

* np.array(), np.arange(),np.dot()
* X.reshape(), X.T

|  |
| --- |
| Code |
| X = np.arange(1,13).reshape((3,4))  w = np.array([0.1,0.2,0.3])  y = w.dot(X)  y |
| Result(Captured images) |
|  |
| Description |
| X를 생성할 때 np.arange와 reshape를 사용해서 shape가 (3,4)인 matrix생성  W는 np.array를 이용해 생성한다. X.shape는 (3,4)이고 w.shape는 (3,)이다. shape가 (4,)인 y를 주어진 식대로 계산하려면 matrix dot product를 이용하면 된다. Dot product를 하려면 행과 열을 같게 맞춰줘야 하므로 transpose메소드 T를 사용해 (4,3) x (3,) = (4,)이므로 y.shape와 같다. |

1-3. Pandas : From Boston Housing Price dataset, compute “DIS” column’s count, mean, std. (Week03-Quiz4)

|  |
| --- |
| Code |
| data\_url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/housing/housing.data'  df\_data = pd.read\_csv(data\_url, sep='\s+', header=None)  df\_data.columns = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS',  'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']  print(df\_data["DIS"].count())  print(df\_data["DIS"].mean())  print(df\_data["DIS"].std()) |
| Result(Captured images) |
|  |
| Description |
| Pandas series는 column내 인스턴스 개수를 계산하는 count, 인스턴스들의 평균을 계산하는 mean, 표준편차를 계산하는 std메소드를 가지고 있다. describe메소드는 count, mean, std뿐만 아니라 min,max, 4분위수들까지 계산해준다 |

1-4. Matplotlib : Plot for with red triangles.

|  |
| --- |
| Code |
| t = np.arange(0, 10, 0.5)  y = 2 \* t\*\*3 + 500  plt.plot(t, y, "r^")  plt.show() |
| Result(Captured images) |
|  |
| Description |
| Np.arange를 이용해서 0~9.5 데이터를 쉽게 생성할 수 있다. 이후, 주어진 식대로 y를 t로 표현한 후, plt.plot의 첫번째 파라미터로 t, 두번째 파라미터로 y를 준 후, 세번째 파라미터로 r^를 주면 알아서 차트가 그려진다. |

2. Explain what Supervised Learning, Unsupervised Learning, and Reinforcement Learning are, and describe the differences. (10 pts)

|  |
| --- |
| Your Answer |
|  |

3. Describe the concept of “overfitting”, and explain how you can prevent overfitting in supervised learning. (20 pts)

|  |
| --- |
| Your Answer |
|  |

4. Describe the differences between Gradient Descent and Stochastic Gradient Decent in detail and explain pros and cons (you can explain by using examples). (20 pts)

|  |
| --- |
| Your Answer |
|  |

5. The seeds.csv dataset represents 7 geometric parameters of wheat kernels for 3 different varieties of wheat. Preprocess the dataset properly and output the cost function graph when you perform AdalineGD and AdalineSGD respectively (Specify the hyperparameter – *η, epoch*, etc.). (30 pts)

|  |
| --- |
| Code |
|  |
| Result(Captured images) |
|  |
| Description |
|  |

**Note**

1. Submit the file to e-class as pdf

2. Specify your pdf file name as “hw1\_<StudentID>\_<Name>.pdf”

Ex) hw1\_2000123456\_홍길동.pdf