**개별 프로젝트 과제 (팀 프로젝트 대체)**

제출기한 : 23년 12월 23일 (토) 23:59분

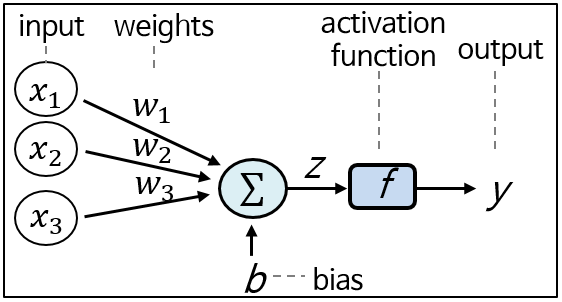
작성한 프로그램 코드 및 실행 결과물을 포함하여 word로 작성해주세요. (**레포트 반드시 첨부, 누락시 0점**)  
**작성한 Programming 코드 (.py 또는 .ipynb) 를 꼭 함께 제출 하세요.  
제출 code에 단순 warning이 아닌 수정 불가능한 error 발생 시 채점 되지 않습니다.**본인이 작성한 코드에 대해서는 주석을 통해 설명을 달아주세요.  
파일이 2개 이상 일 경우 하나의 .zip파일로 업로드 하세요.   
**＃ Copy 엄금, 적발 시 0점, 딜레이는 받지 않습니다 (기간 엄수)**

**[참고 1]**데이터 로드 및 평면 출력

|  |
| --- |
| %matplotlib notebook  from mpl\_toolkits.mplot3d import Axes3D  import matplotlib.pyplot as plt  from matplotlib import cm  import numpy as np  import pandas as pd  data = pd.read\_csv("p1\_training\_data.csv") # 데이터 읽기  np\_data = np.array(data)  fig = plt.figure()  ax = fig.add\_subplot(111, projection='3d')  ## positive samples  x\_1 = np\_data[0:50,0]  y\_1 = np\_data[0:50,1]  z\_1 = np\_data[0:50,2]  ## negative samples  x\_0 = np\_data[50:,0]  y\_0 = np\_data[50:,1]  z\_0 = np\_data[50:,2]  ## Generation 1의 fittest gene  w1 = 0.28645574  w2 = -0.43628723  w3 = 0.30481866  b = -14.39337271  ax.plot(x\_1, y\_1, z\_1, linestyle="none", marker="o", mfc="none", markeredgecolor="b") #샘플 출력  ax.plot(x\_0, y\_0, z\_0, linestyle="none", marker="o", mfc="none", markeredgecolor="r") #샘플 출력  X = np.arange(0, 2, 0.1)\*100  Y = np.arange(0, 2, 0.1)\*100  X, Y = np.meshgrid(X, Y)  Z = (-float(w1)/w3 \* X) + (-float(w2)/w3 \* Y) - float(b)/w3 # 평면의 방정식  ax.plot\_surface(X, Y, Z, rstride=4, cstride=4, alpha=0.4, cmap=cm.Blues) # 평면 출력  plt.show() |

**Genetic Algorithm**

**1 유전 알고리즘을 통해 주어진 데이터를 선형 분류하는 single-layer neural network를 학습하라. [전체: 50점]**아래 세부 사항을 확인하여 문제를 해결 하시오.

* 학습 데이터: p1\_training\_data.csv (100개x4차원데이터, 1~3차원 데이터는 입력데이터 (x1,x2,x3), 4번째 데이터는 클래스 y(1 또는 0))
* 학습 파라미터: **w** (3차원), b (1차원) [아래 그림 참조]  
  
* Activation function: Sigmoid function 을 사용할 것.
* Population수는 100 이하로 설정하라.
* 위 조건 이외의 유전알고리즘 동작을 위한 모든 설계 조건은 자율적으로 설계 할 것 (**단, 레포트에 설계 내용을 모두 기입**)
* (fitness계산방법, Selection방법, crossover방법, mutation방법, 유전자(학습 파라미터)초기화 방법, 알고리즘 종료 조건 등)
  1. 1세대 유전자를 초기화 하라. 어떤 방식으로 유전자를 초기화 하였는가? Population은 얼마로 세팅하였는가? [5점]  
     (hint. 각각의 weight는 -1~1사이로, bias는 -100~100 사이로 초기화 한다.)

genes = [np.random.uniform(-1, 1, size=100) for \_ in range(3)] 구문을 이용하여 각각 population이 100인 weight 3종류를 만들었다. 또한 bias = np.random.uniform(-100, 100, size = 100) 구문을 이용하여 population이 100인 1차원 배열 형태의 bias를 구현하였다.

* 1. 각 유전자에 대한 fitness 계산법을 설계하고 가장 fitness가 높은 유전자들을 선별하라. [10점]  
     fitness 계산은 어떻게 하였는가?   
     1세대에서 가장 fitness가 높은 상위 4개의 유전자에 대해 fitness score를 적고, 분류 평면을 도시화 하라 [실행 예 참고]

|  |  |  |  |
| --- | --- | --- | --- |
| Ranking1, Fitness score: 30.48 | Ranking2, Fitness score: 36.14 | Ranking3, Fitness score: 44.03 | Ranking4, Fitness score: 45.89 |

표1. [실행 예] 1세대 유전자 (Rank1~4)

먼저 sigmoid 함수와 zeta 함수를 구현하여 가중치에 따른 결과값을 0부터 1 사이의 값으로 표현하였다. 이후, fitness 함수를 구현하여 weights와 bias에 따른 fitness값들을 반환하게 하였다. fitness값을 구할 때는 p1\_training\_data.csv 파일 속 positive samples와 negative samples, weights, bias를 활용하여 zeta 함수와 sigmoid 함수를 적용시켜 각각 y1, y0의 값으로 지정한 후, 구한 값들을 fit에 각각 저장하는 방식을 선택했다.

* 1. 선택된 유전자들에 대해 Crossover와 mutation을 수행하여 2세대 유전자들을 생성하시오. 이때 생성된 자식세대의 population 수는   
     1세대와 동일하게 설계 하시오.  
       
     Crossover는 어떻게 수행하였는가? 상세히 작성하시오. [10점]  
       
     중첩 반복문의 매개변수 I, j를 이용하여 각 genes와 bias의 i \* 4 + j번째 값을 rank\_genes에 저장된 우수한 개체로 변경하였다. 이때, crossover를 위해 genes[0], genes[1]은 rank\_genes의 i번째에 해당하는 값으로, genes[2], bias는 rank\_genes의 j번째에 해당하는 값으로 각각 지정해주었다.   
       
     Mutation은 어떻게 수행하였는가? 상세히 작성하시오. [10점]

crossover 과정으로 인해 각 genes의 0 ~ 15번째 인덱스까지는 우수한 개체들이 저장되어 있는 상태이다. 그렇기에 for문을 84번 돌리고, 매개변수 i를 이용하여 genes와 bias의 I + 16번째 값들을 우수한 개체에 0.5부터 1.5까지의 랜덤한 값을 곱해주어 변이를 일으켰다.

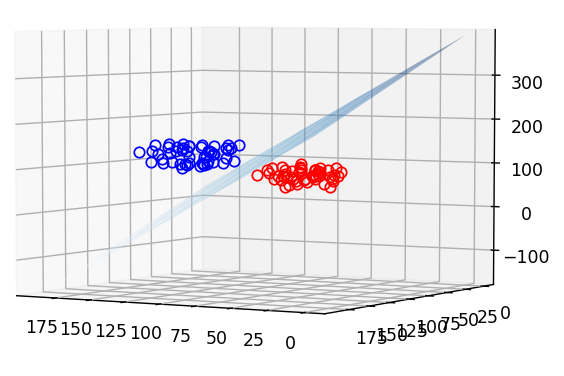
* 1. (2,3)과정을 반복하여 2세대, 3세대에서 가장 fitness가 높은 상위 4개의 유전자에 대해 fitness score를 적고, 분류 평면을 도시화 하라. [10점]

|  |  |  |  |
| --- | --- | --- | --- |
| Ranking1, Fitness score: | Ranking2, Fitness score: | Ranking3, Fitness score: | Ranking4, Fitness score: |

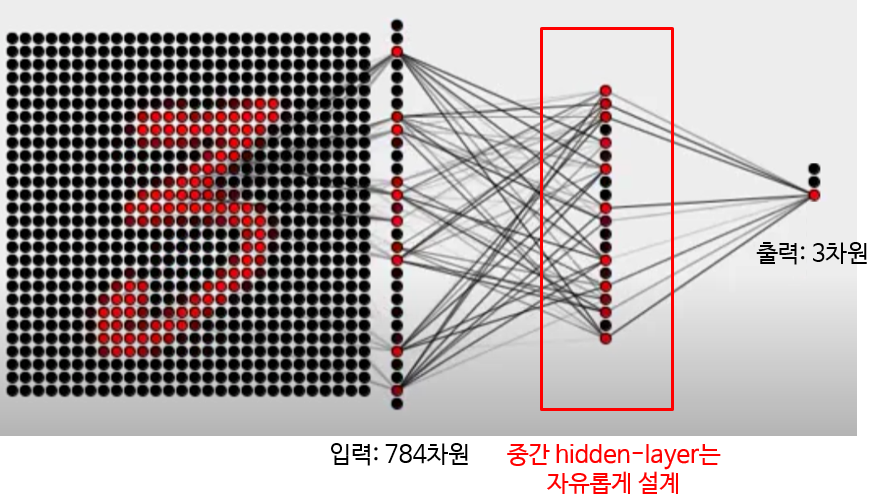
표2. 2세대 유전자 (Rank1~4)

|  |  |  |  |
| --- | --- | --- | --- |
| Ranking1, Fitness score: | Ranking2, Fitness score: | Ranking3, Fitness score: | Ranking4, Fitness score: |

표3. 3세대 유전자 (Rank1~4)

* 1. 최종적으로 유전알고리즘을 통해 얻어진 유전자중 가장 fitness가 높은 유전 자에 대해 분류 평면을 도시화 하라. 또한 어떤 조건으로 유전 알고리즘을 종료 하였는지 작성하라. [5점]  
     fitness 함수로 나온 결과값 중 최소값이 0.0000001보다 작아질 경우, 즉 0에 가까워질 경우 유전 알고리즘을 종료하도록 했다.   
     [실행예]  
     3세대 fittest gene. Fitness score : 0.000012  
     

**2. 유전 알고리즘을 통해 주어진 데이터를 분류하는 Multi-layer neural network를 학습하라. [전체: 50점]**아래 세부 사항을 확인하여 문제를 해결 하시오.

* [학습 데이터]  
  이미지 데이터: p2\_training\_data\_mnist.csv (600개x784차원데이터)  
  1~3 숫자 600개데이터가 저장 되어있다. 하나의 숫자는 28x28=784차원의 데이터이다.  
    
  Label 데이터: p2\_training\_data\_label.csv (600개x3차원데이터)  
  3차원으로 표현되며 숫자 1은 [1,0,0], 2는 [0,1,0], 3은 [0,0,1]의 형태로 저장 되어있다.
* 숫자 이미지 1,2,3을 분류할 수 있는   
  Multi-layer neural network의 구조는 자유롭게 설계할 것.   
  (단, 입력은 784차원, 출력은 3차원으로 설계. 중간 layer의 node 수 및 층 수는 자유롭게 설계)  
  [아래 그림 참조]  
  
* Activation function: Sigmoid function 을 사용할 것.
* 유전알고리즘 동작을 위한 모든 설계 조건은 자율적으로 설계 할 것.
* (Population수, fitness계산방법, Selection방법, crossover방법, mutation방법, 유전자(학습 파라미터)초기화 방법, 알고리즘 종료 조건 등)

2-1) 설계한 네트워크의 구조는 어떻게 되는가? (입력과 출력층 사이에 적어도 1개 이상의 층을 반드시 설계하라.) [10점]  
 (ex. Layer1(input): 784nodes / Layer2: 18nodes / Layer3(output): 3nodes)

2-2) 유전알고리즘을 통해 2-1)에서 설계한 네트워크를 학습하시오. 알고리즘이 종료될때까지, 각 세대별로 가장 높은 fitness 를 출력하시오. [20점]  
[실행 예]  
Generation 1 best fitness : 580.3221  
Generation 2 best fitness : 215.1078  
Generation 3 best fitness : 126.6617  
Generation 4 best fitness : 150.3162  
…  
Generation 10 best fitness : 0.0122

2-3) 학습한 최종 네트워크(가장 높은 fitness)로 p2\_test\_data\_mnist.csv (30개x784차원데이터) 데이터에 대한 test를 수행하라. [20점]  
[실행 예]  
Test 1 : [1, 0, 0]  
Test 2 : [1, 0, 0]  
Test 3 : [1, 0, 0]  
…  
Test 30 : [0, 0, 1]  
  
[참고]  
Test 결과는 아래 예시와 같이 0~1사이의 확률 값으로 출력이 될 수 있다.   
예시) [0.991, 0.011, 0.021] 이 경우 아래와 같이 값이 가장 큰 결과를 1로, 나머지는 0으로 변환하여 최종 결과를 얻는다.  
예시) 네트워크 출력 값: [0.991, 0.011, 0.021] → 최종 숫자 판별 값 [1, 0, 0]