

인공지능 과제 리포트

과제 제목: K-Nearest Neighbor

학번: B811226

이름: 이현동

1. 과제 개요

과제의 주제와 과제가 어떤 내용을 다루는지 간단하게 설명

K-nearest neighbor를 직접 코드로 구현한다. k-nearest neighbor를 classification에 이용한다. 사이킷 런에서 제공하는 iris_data를 이용하여 꽃을 분류한다. Knn class를 구현하여 main 함수에서 KNN class로부터 각각의 k=3, 5, 10일 때를 출력하여 예상했던 출력값과 실제 출력값을 비교한다.

2. 구현 환경

Window 내 wsl ubuntu LTS -v 20.xx

Anaconda/Jupyter notebook/envs 이용.

이후 .py 파일 확장자를 이용하기 위해 Wsl(anaconda envs) + VScode 이용.

3. 알고리즘에 대한 설명

K-nearest neighbor는 classification을 할 때 사용된다. distance measurement를 이용해서 동작을 시키는데 k를 지정하여 정해진 이웃의 개수로 묶어서 k개의 근접한 데이터 중에서 가장 많은 비중의 클래스로 분류하는 방식이다. 이때 사용하는 voting은 크게 두 가지로 나뉘는데 majority vote와 weighted majority vote가 있다.

4. 데이터에 대한 설명

4.1 Input Feature

Input feature의 차수는 (150, 4) 150x4이고 feature의 종류로는 Sepal length(cm), sepal width(cm), petal length(cm), petal width(cm) 네 가지의 feature가 있다. Data의 타입은 array이다. iris data 값에는 array 형태로 feature가 mapping되어 있다.

4.2 Target Output

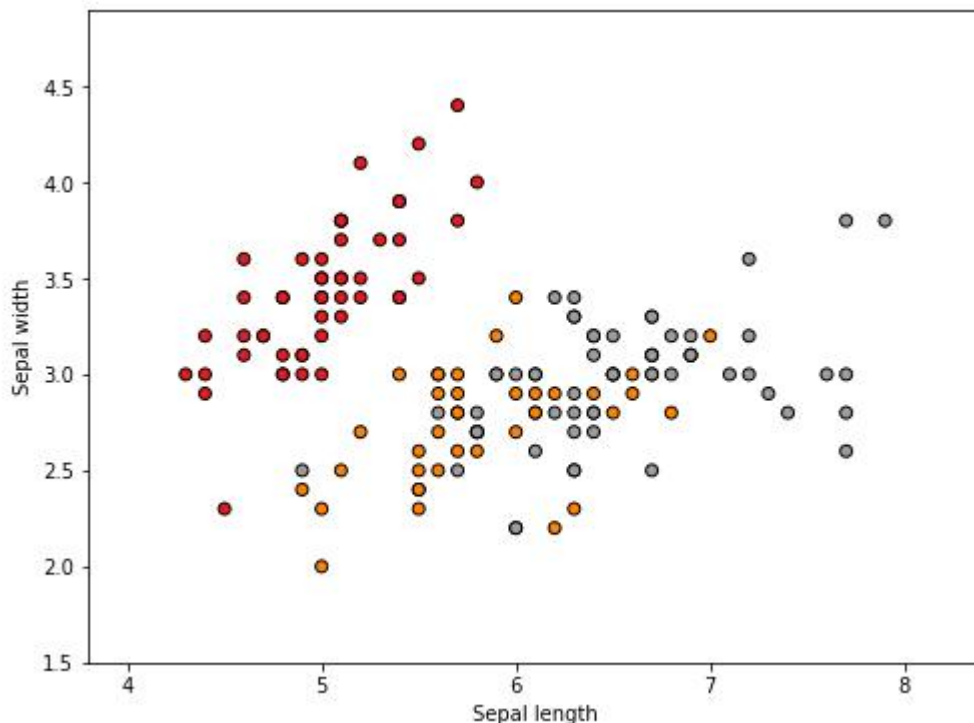
target의 값은 0, 1, 2가 있다. 즉 3가지 클래스를 가진다. target_name은 ['setosa', 'versicolor', 'virginica']가 있다. (150,) 즉 150x1 차원이다.

5. 소스코드에 대한 설명

우선 split_data 부분을 구현하기 위해서 기존에 split_data함수를 사용하지 않고 직접 구현을 했다. X, y의 각각 train, test 값을 14:1 비율로 나눠주는데 특이점은 각 순서의 15번째 데이터를 test data로 넣어줘야 나름 균등하게 class값이 나뉜다. 그래서 for문 O(n)의 시간복잡도로 15번째 데이터를 조건문으로 나눠서 넣어줬다. 또 weightedMajorityVote를 이용하기 위해서 weight값을 지정해줘야 하는데 $1/(1+d)$ 나 $\exp(-d)$ (d는 거리) 이런 단조 감소 함수를 이용하여 작성하는데 나

의 코드에서는 $1/(1+d)$ 이용하여 weight로 각 distance에 곱하여 weight를 달아줬다.

6. 학습 과정에 대한 설명



두 가지의 feature만 가지고 plotting을 하였다. 3차원으로 plotting을 하면 시각적으로 더 보기 불편하기 때문에 위 그림을 이용한다. 위 그림처럼 근접한 이웃들이 각각의 점마다 존재한다. 우선 기준점이 데이터 개수만큼 존재한다. 예를 들어서 X의 첫 번째 인덱스와 근접한 이웃을 찾는다고 할 때 나머지와 거리를 계산해야한다. 즉 distance matrix에 0이 되는 값은 자기 자신이고 나머지는 거리를 euclidean distance를 계산해준다. 각각의 test data에서 train data까지의 distance를 계산하고 plot(한 점)과 가장 가까운 순서대로 sort를 해야 한다. 우리는 가장 가까운 거리를 sorting한 함수가 필요한 게 아니라 가까운 거리 순의 index의 값이 필요하다. 왜냐하면 class의 값을 출력해 줘야하기 때문이다. Numpy의 argsort를 이용하여 거리가 가까운 순서대로의 index값을 저장한 matrix를 리턴 받는다. 그 뒤에 voting을 진행한다. k의 값에 따라서 가장 근접한 값들의 비중을 차지하는 클래스가 무엇인지 majorityVote함수와 WeightedMajorityVote를 이용하여 class를 리턴 받아서 마지막으로 knn함수를 통해서 class의 이름(target_name)을 리턴해준다.

7. 결과 및 분석

결과 출력 및 결과에 대한 분석

using majority vote

k == 3

Test Data Index: 0 Computed class: setosa , True class : setosa

Test Data Index: 1 Computed class: setosa , True class : setosa

Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: versicolor , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica

k == 5

Test Data Index: 0 Computed class: setosa , True class : setosa
Test Data Index: 1 Computed class: setosa , True class : setosa
Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: versicolor , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica

k == 10

Test Data Index: 0 Computed class: setosa , True class : setosa
Test Data Index: 1 Computed class: setosa , True class : setosa
Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: virginica , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica

using weighted majority vote

k == 3

Test Data Index: 0 Computed class: setosa , True class : setosa
Test Data Index: 1 Computed class: setosa , True class : setosa
Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor

```
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: versicolor , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica
```

```
k == 5
```

```
Test Data Index: 0 Computed class: setosa , True class : setosa
Test Data Index: 1 Computed class: setosa , True class : setosa
Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: versicolor , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica
```

```
k == 10
```

```
Test Data Index: 0 Computed class: setosa , True class : setosa
Test Data Index: 1 Computed class: setosa , True class : setosa
Test Data Index: 2 Computed class: setosa , True class : setosa
Test Data Index: 3 Computed class: versicolor , True class : versicolor
Test Data Index: 4 Computed class: versicolor , True class : versicolor
Test Data Index: 5 Computed class: versicolor , True class : versicolor
Test Data Index: 6 Computed class: virginica , True class : virginica
Test Data Index: 7 Computed class: virginica , True class : virginica
Test Data Index: 8 Computed class: virginica , True class : virginica
Test Data Index: 9 Computed class: virginica , True class : virginica
```

weighted majority vote나 majority vote 모두 같은 출력 값을 얻게 된다. 하지만 근접한 이웃의 사이즈가 너무 작을 때는 제대로 classify하지 못하는 상황도 발생한다. 모두 k=3, 5일 때는 잘못된 분류를 한 개씩 했고 이외에 값 k =10일 때 모두 true class와 같은 값을 출력하는 것을 확인할 수 있다.