인공지능

과제 제목: KNN with MNIST Data

이름: 박수민

1. 개요

-K-Nearest Neighbor 알고리즘을 이용하여 mnist data 분류하기

1. 구현 환경

-Pycharm

1. 알고리즘에 대한 설명

-K-NN 알고리즘은 머신러닝 알고리즘 중 지도학습에 속한다.

-시간 복잡도: O(n)

-거리 측정 방법: Euclidean distance

텍스트이(가) 표시된 사진

자동 생성된 설명

-Number of neighbors to consult: k

-Majority vote: k개의 가까운 데이터 중에서 가장 수가 많은 데이터를 major로 선정.

-Weighted majority vote: k개의 가까운 데이터 중에서 거리에 따르는 가중치를 합한 값에서 major를 선정.

-장점: 구현과 이해가 쉽다. Training이 빠르다, 정확도 높다

-단점: 계산양이 많다. O(n)만큼 저장공간. k값에 sensitive.

1. 데이터에 대한 설명

* Input Feature
* Data Dimension: 784 (28x28)
* Data Type: images
* Training Data Size: 60,000
* Test Data Size: 10,000

1. 소스코드에 대한 설명

5.1

텍스트이(가) 표시된 사진

자동 생성된 설명

-테스트케이스와 나머지 데이터들의 거리를 Euclidean방식으로 구현함

5.2

텍스트이(가) 표시된 사진

자동 생성된 설명

-k개의 이웃한 neighbor를 구한 뒤 sort\_target에 k개의 값들을 정렬함

5.3

텍스트이(가) 표시된 사진

자동 생성된 설명

-list\_target에는 k개의 이웃한 neighbor가 정렬되어 있는데, counter함수와, most\_common함수를 이용해 가장 많은 값을 추출하는 majority vote를 구현함

5.4

텍스트이(가) 표시된 사진

자동 생성된 설명

-list\_target에 정렬되어 있는 순서가 가까운 거리부터 오름차순이므로 인덱스 작은 순서부터 가중치를 더한 뒤 가중치가 가장 높은 데이터를 찾을 수 있도록 하였음

1. 학습 과정에 대한 설명

-손 글씨 데이터 예시-

화살이(가) 표시된 사진

자동 생성된 설명

-데이터 배열 예시-

테이블이(가) 표시된 사진

자동 생성된 설명

Train Data 값과 Test Data 값을 28x28 총 784개의 input으로 거리를 구한 뒤 가까운 k개 중에서 가중치를 더한 결과값으로 데이터를 예측하도록 했습니다.

1. 결과 및 분석

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

* Accuracy가 0.67으로, 약 67%의 정확도를 보였습니다.