

■ 설명

● MNIST 데이터

LABEL 파일의 첫 4바이트는 magic number라 명시된 32비트 정수이며, 다음 4바이트는 항목의 개수를 32비트 정수로 나타낸다. 이후 내용은 1바이트마다 비부호 바이트형을 가지며 이는 0 ~ 9의 범위를 가지는 분류값을 나타낸다.

이 파일에서는 클래스의 개수를 알 수 있으며, 0 ~ 9로 분류되므로 총 10개이다..

IMAGE 파일의 첫 4바이트는 magic number라 명시된 32비트 정수이며, 다음 4바이트는 이미지의 개수를 32비트 정수로 나타낸다. 다음 8바이트는 이미지마다 가지는 행의 수, 열의 수를 32비트 정수로 각각 나타낸다. 이후 내용은 1바이트마다 비부호 바이트형을 가지며 이는 0 ~ 255의 범위를 가지는 픽셀값을 나타낸다.

이 파일에서는 특징값의 개수를 알 수 있으며, 이미지 픽셀의 개수가 $28 * 28 = 784$ 개이므로 총 784개이다.

이 포맷에서는 magic number(정수) -> 0차원 요소의 개수(정수) -> 1차원 요소의 개수 -> ..., N차원 요소의 개수 -> 이후 데이터 순으로 내용이 존재하며 이는 Big Endian 형태로 저장되어 있다.

● 비교 방법

전체 데이터를 학습 데이터와 실험 데이터로 나누었고, 시간이 오래 걸림에 따라 비율을 0.05 : 0.95로 하였다.

SVM 모델에 학습시킨 후에, 맞게 예측한 결과의 수를 실험 데이터의 수로 나누었다.

■ 실행결과 및 비교

[파일 정보]

Image 개수 : 60000

Image 당 행 개수 : 28

Image 당 열 개수 : 28

Label 개수 : 60000

[데이터셋 구성 완료]

Image Data 개수 : 60000

label Data 개수 : 60000

[학습데이터, 실험데이터 분리 완료]

X_train 개수 : 3000

X_test 개수 : 57000

y_train 개수 : 3000

y_test 개수 : 57000

[svm setting complete]

[linear Training complete]

[nonlinear Training complete]

[linear Prediction complete]

[nonlinear Prediction complete]

Linear SVM 정확도 : 0.801912

nonLinear SVM 정확도 : 0.112263

Process finished with exit code 0

Linear SVM이 더 정확도가 높게 나왔다.