

Introduction to Machine Learning

Mini-project 1: Radial Basis Function Networks

Problem Overview:

Radial basis function network를 구현한다. 특별히 본 모델에 필요한 kernel function을 설계해야 한다. 구현된 분류기를 바탕으로 MNIST 데이터를 분류한다.

Problem Details:

프로그래밍 언어로 Python 2.7을 사용하되, 기본 라이브러리 및 NumPy 라이브러리 사용을 허용하고, Theano 및 기타 라이브러리 사용은 불허한다. 실행 환경의 디렉토리 구조는 다음을 가정한다.

```
code_학번명/  
    run_rbf.py  
    (이외 다른 소스 파일)  
    result.txt  
data/  
    mnist.pkl.gz
```

code_학번명(예, code_2015-12345)으로 된 디렉토리에 작성한 모든 소스 파일을 담는다. run_rbf.py 이라는 script file을 실행하면 같은 디렉토리 안에 result.txt 라는 파일에 첫 2천 개의 테스트 데이터에 대해 분류한 label(0-9)을 띄어쓰기(space)로 구분하여 저장하는 프로그램을 만든다. 이 때, 실행 코드가 있는 code_학번명 디렉토리와 같은 레벨에 data 디렉토리가 있고, 그 안에 MNIST 데이터 셋 mnist.pkl.gz이 있는 상황을 가정한다. 이 MNIST 데이터 셋의 형식과 이름은

<http://deeplearning.net/data/mnist/mnist.pkl.gz>

에 제공된 파일과 같다. 구현된 프로그램은 NumPy가 설치된 기본적인 Linux 환경에서 가동될 것이다.

학습과 추론에 10분 이상 걸리는 프로그램은 동작하지 않는 것으로 간주한다.

Submission:

10월 8일 23:59:59까지 chhan@bi.snu.ac.kr 로 code와 보고서를 zip파일로 묶어서 제출한다. 메일 제목, zip 파일, 보고서에 과목 이름과 학번 그리고 이름을 명시한다 (e.g. [기계학습개론] 2015-12345 홍길동). 제출되는 code 안에 MNIST 데이터 셋을 포함하지 않는다. 제출 형식을 지키지 않을 경우 감점한다.

Report:

보고서는 폰트 10, A4 5페이지 이내로 작성하는 것을 권장한다. 보고서의 서술 방법은 자유롭게 정할 수 있으나 다음 내용을 포함해야 한다.

- 구현된 모델에 대한 서술
- 성능 보고
- Gaussian kernel의 변화에 따른 비교

Data:

실습 때 사용된 MNIST Dataset이 평가에 사용될 데이터이다. 학습 데이터로 MNIST 학습 데이터의 첫 10000개를, 테스트 데이터로 MNIST 테스트 데이터의 첫 2000개를 사용하는 것을 기본데이터 셋으로 한다. 위에 언급된 시간 제한은 이 데이터 셋에 대한 것이다. 보고서의 성능 평가 역시 이 기본 데이터 셋에 대한 성능 보고를 포함해야 한다.

Grading:

본 프로젝트의 평가요소는 보고서 (60%), 성능(30%), 기타 형식(10%)이다. 보고서에서는 구현된 알고리즘에 대한 이해 정도, 그리고 구현 및 실험에 대한 설명의 명확성 등을 평가한다. 성능 측면에서는 구현된 알고리즘이 주어진 테스트 데이터를 얼마나 잘 예측하는 지를 본다. 이때, 프로그래밍 언어로 MATLAB을 선택한 경우와 구분하여 평가한다. 별도의 테스트 데이터 셋을 두고 프로그램을 실행하여 채점할 것이다. 학습과 추론의 속도는 10분 이상 걸릴 경우 동작하지 않는 것으로 판단하고, 빠른 경우에 대한 가산점은 없다. 기타 형식에서는 파일 디렉토리 구조, 메일 형식 등이 틀릴 경우 감점한다.

Penalty:

Cheating에는 강한 패널티가 부여된다. 타인의 코드나 최종 보고서를 공유하는 것은 cheating에 해당된다. 기본 라이브러리가 아닌 Theano 및 기타 라이브러리 사용 역시 cheating에 해당된다.

Delay에는 24시간마다 매번 10%씩의 Penalty가 적용된다. 이를 Delay 된 경우 Delay하지 않은 경우와 비교하여 81%의 점수를 받을 수 있다.