

Deep Learning Lecture 1 - Examples overview

MA8701 General Statistical Methods

Thiago G. Martins, Department of Mathematical Sciences, NTNU

Spring 2019

- MNIST dataset
 - Objective
 - Data type
 - Problem type
- IMDB dataset
 - Objective
 - Data type
 - Problem type
- Reuters dataset
 - Objective
 - Data type
 - Problem type
- The Boston housing price dataset
 - Objective
 - Data type
 - Problem type
- References

This page provides a simple overview of the examples covered in the lectures. This lecture note is based on (Chollet and Allaire 2018).

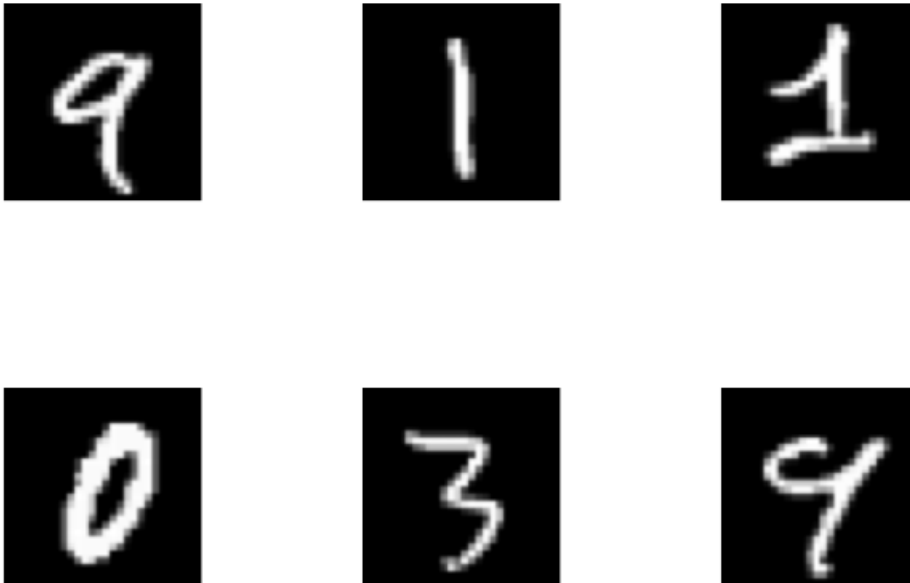
MNIST dataset

Objective

- classify the digit contained in a image.

Data type

- We have images such as



- with annotated labels

```
## train_labels
##      0      1      2      3      4      5      6      7      8      9
## 5923 6742 5958 6131 5842 5421 5918 6265 5851 5949
```

Problem type

- Multiclass classification based on image data.

IMDB dataset

Objective

- classify a movie review as either positive or negative.

Data type

- We have text data

```
## [1] "? this film was just brilliant casting location scenery story direction
```

- with annotated labels where 0 stands for negative and 1 stands for positive.

```
## train_labels
##      0      1
## 12500 12500
```

Problem type

- binary classification based on text data.

Reuters dataset

Objective

- classify short news stories into one of 46 topics available.

Data type

- We have text data

```
## [1] "? ? ? said as a result of its december acquisition of space co it expect
```

- with annotated labels going from 0 to 45

```
## train_labels
##      0      1      2      3      4      5      6      7      8      9     10     11     12     13     14
##  55  432   74 3159 1949   17   48   16  139  101  124  390   49  172   26
##  15   16   17   18   19   20   21   22   23   24   25   26   27   28   29
##  20  444   39   66  549  269  100   15   41   62   92   24   15   48   19
##  30   31   32   33   34   35   36   37   38   39   40   41   42   43   44
##  45   39   32   11   50   10   49   19   19   24   36   30   13   21   12
##    45
##    18
```

Problem type

- multiclass classification based on text data.

The Boston housing price dataset

Objective

- Predict the median price of homes in a given Boston suburb in the mid-1970s, given data points about the suburb at the time, such as the crime rate, the local property tax rate, and so on.

Data type

- We have a small dataset and 13 numerical features

```
## num [1:404, 1:13] 1.2325 0.0218 4.8982 0.0396 3.6931 ...
```

- The targets are the median values of owner-occupied homes, in thousands of dollars:

```
## num [1:404(1d)] 15.2 42.3 50 21.1 17.7 18.5 11.3 15.6 15.6 14.4 ...
```

Problem type

- regression based on numerical features.

References

Chollet, F., and J. Allaire. 2018. *Deep Learning with R*. Manning Publications.
<https://books.google.no/books?id=xnIRtAEACAAJ>.