A study on sentiment classification of news headlines using deep learning

CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

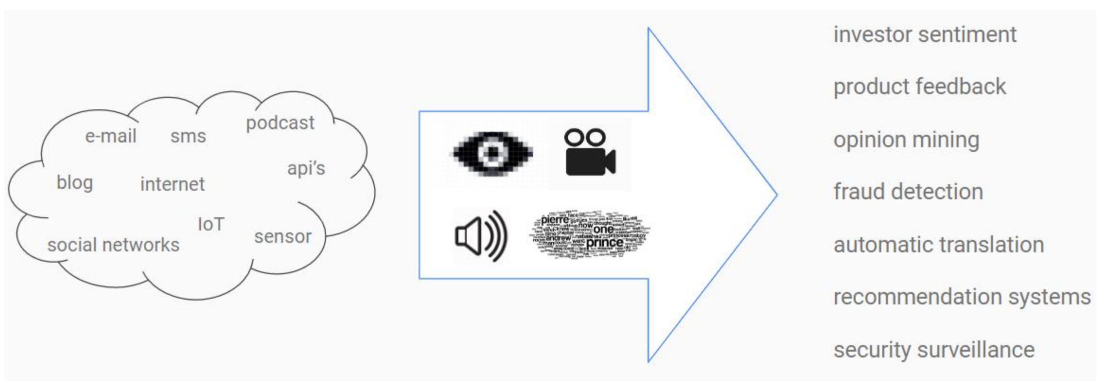
Higher Diploma in Science in Data Science & Analytics – Level 8
Department of Mathematics

Joao Tiago Viegas
Supervisor: Dr. Laura Climent

Introduction and Methodology

Motivation

In the latest years we have experienced an explosion of data in our daily lives. Short-Message-Service(SMS), the internet, the social networks, blogs, posts, podcasts, smartphones, the Internet-of-Things (IOT) and the extensive availability of sensors and System-On-Chip solutions, it feels as if this phenomenon is only starting yet. This data is out there for everyone to tap into and leverage in ways that we are still discovering nowadays. Machine learning (ML) has been extensively used to process this data, and one example is solving natural language processing (NLP) tasks, as for instance, opinion mining.



Recently, a subfield of ML called Deep Learning (DL) has been playing a prominent role in solving these kind of NLP tasks. The main motivation of this study is then to apply and assess some DL models to sentiment classification of news headlines, using a classical classification ML model as baseline: Naïve Bayes.

Dataset

The data we used in this study are the news headlines retrieved from the US Reuters site ranging from January, 8th 2007 through October, 2nd 2018.

	A	B
585	Microsoft, Oracle databases gain share vs	−1
586	Chordiant and Help Global Brands DAK and ING Strengthen Customer Relationships	1
587	signs 10-year multi-billion cloud deal with ABN Amro	1

The data was then classified manually according to its perceived sentiment, using three classes: positive(1), neutral(0) and negative(-1).

Naïve Bayes – baseline model

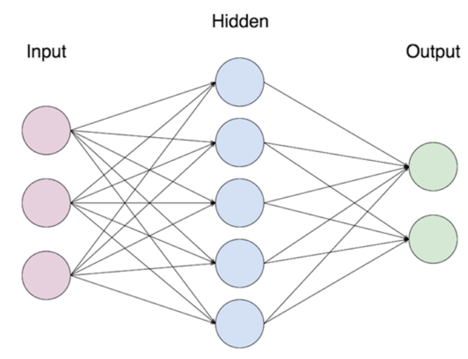
Naïve Bayes is one of the most simple machine learning classification algorithms. It's called naïve because its formulation makes some naïve assumptions, mainly that the probability distribution of its features is independent from the other features distribution. It is based on conditional probabilities, specifically on the Bayes rule. In our study, a multi-class problem with single label, the samples are news headlines and the features translate to words. The Bayes theorem can then be defined

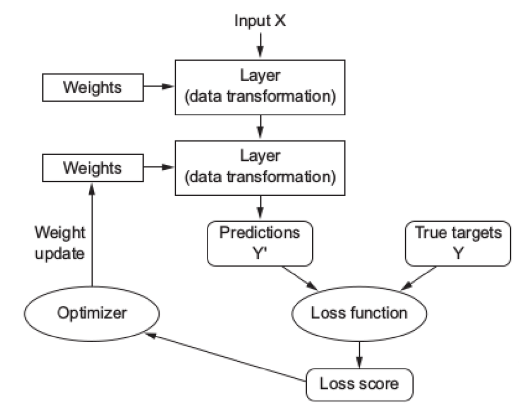
$$P(word|classA) = \frac{P(classA \mid word) * P(word)}{P(classA)}$$

The sum of these conditional probabilities will eventually determine the most likely class of an headline.

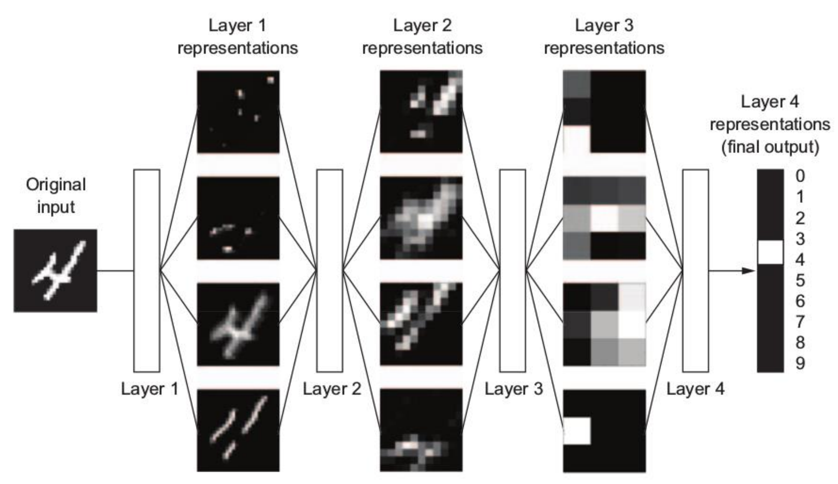
Deep learning – artificial neural networks

Deep Learning (DL) is a subfield of ML. Its rationale is to learn through a sequence of layers of increasingly meaningful representations. The "deep" in the name is related to the number of these successive layers, also called the depth of the model.





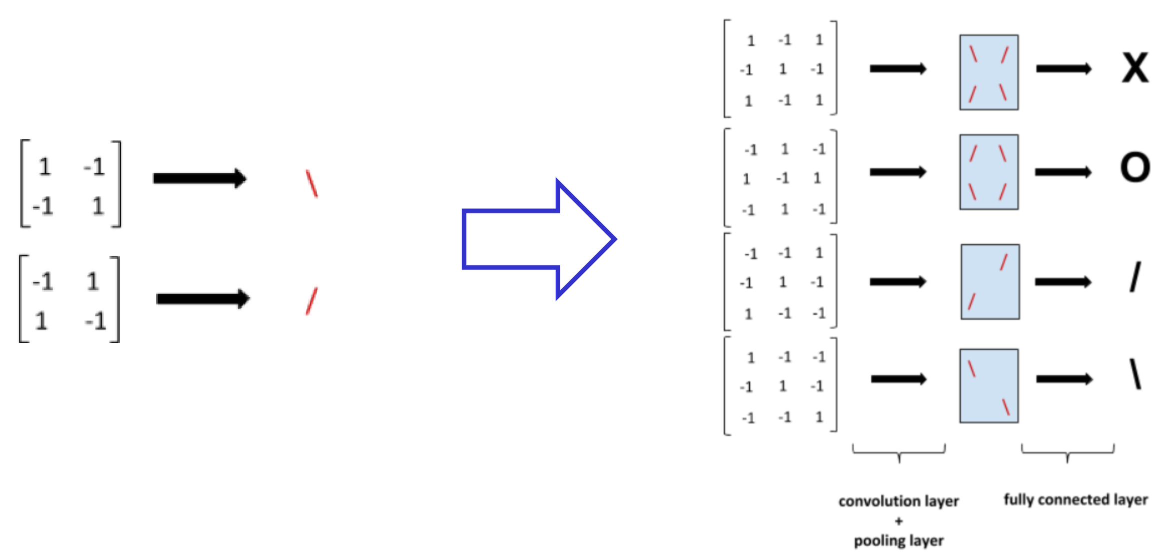
In a simplified manner, we can think of these layers as equations, whose parameters/weights are tuned during the training process, based on the difference on any iteration between their overall value and the desired output (training classification) which is called the loss.



The tuning, performed by the optimizer function, applies a technique called gradient descent, which tries to decrease this loss at every step of the training phase.

Deep learning – convolutional neural networks

Convolutional neural networks are combinations of pairs of convolution+pooling layers and fully, dense connected layers. Firstly the network develops knowledge through training in the form of sub-patterns (filters) with lower dimension.

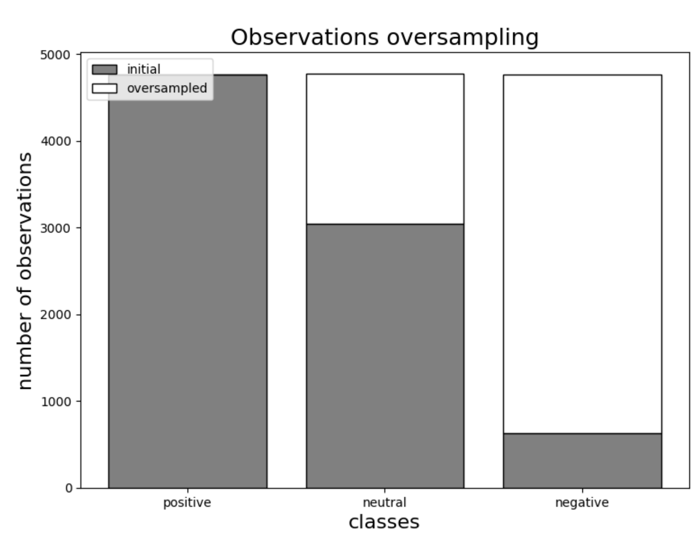


Then the convolutional and the pooling layer split the input data in pieces and try to match it against those known patterns, called filters, creating a kind of a map as output that in turn is used by the fully connected layers to logically relate it with the its classification.

Findings

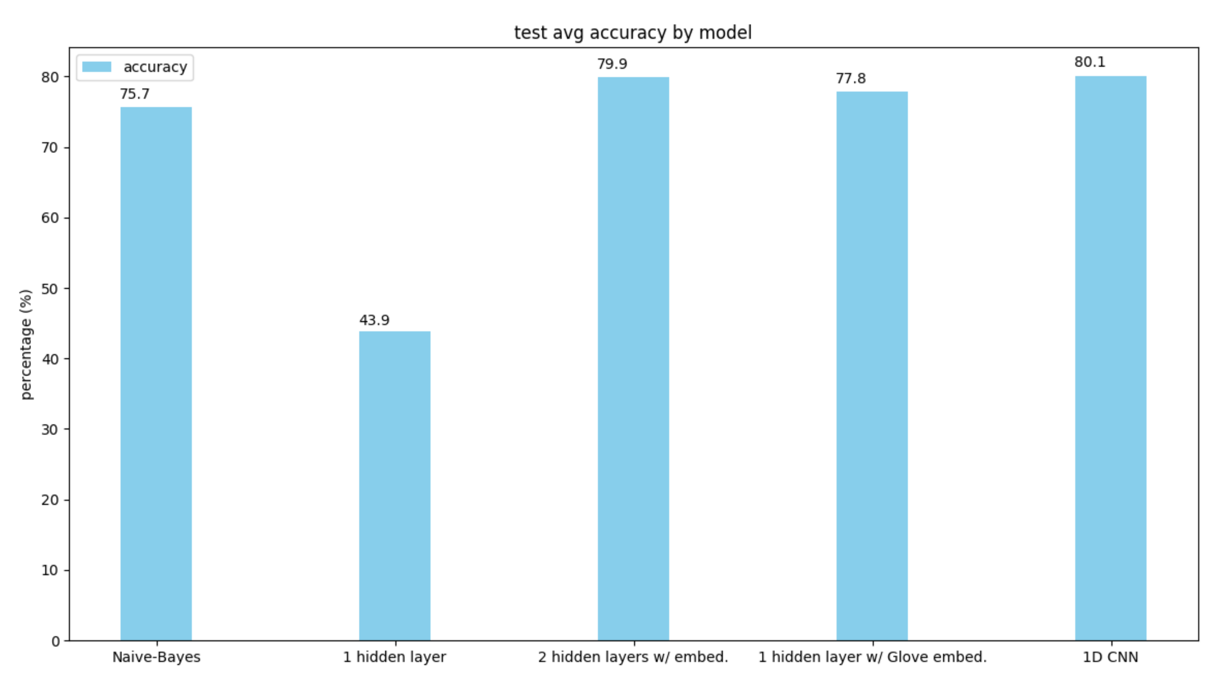
Data preparation

The dataset comprised 9385 observations. The distribution of the dataset classification was imbalanced. Using 10-fold cross validation, training data had 8447 observations. We then oversampled neutral (3044) and negative(632) to the same number of positive observations (4771). We identified and removed punctuation, removed stop words and lemmatized the words. We then shuffled the data at every experiment run.



Models Average Accuracy

Apart from the Naïve Bayes model, we have studied 3 types of artificial neural networks (1 hidden layer, 2 hidden layers with word embeddings, 1 hidden layer with GloVe pre-trained word embeddings) and one convolutional neural network (1D CNN) using the Keras python framework.



Conclusions

- 1D CNN achieved better accuracy;
- 2 layers neural network similar to 1D CNN;
- GloVe pre-trained word embeddings vector not reflecting domain knowledge;
- Naïve Bayes not far away;

Ideas for future work

- include full news article content;
- bigger dataset;
- further parameter testing;
- further topology experimentation;