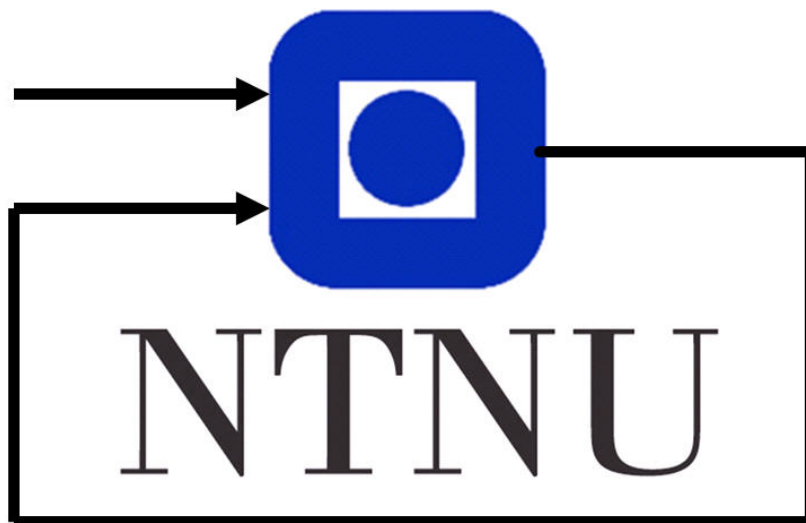


Assignment 5
TDT4171 Artificial Intelligence Methods

ALEKSANDER ELVEBAKK

April 4, 2020



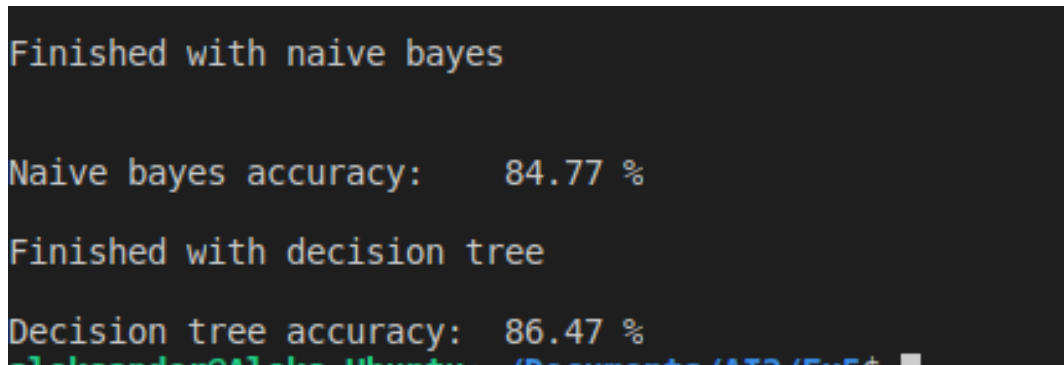
Contents

1	Using sklearn	1
1.1	Naive Bayes	1
1.2	Decision Tree	1
1.3	Using HashingVectorizer	1
2	Using keras	2
2.1	Reasons for good performance	2
2.2	Parameters	2
2.3	Accuracy	2

Introduction

The goal of this exercise is to use several different learning algorithms to classify a review for a restaurant as either positive or negative. I used Python 3.6.9 and scikit-learn 0.22.2 in Part 1 and tensorflow 1.14.0 keras 2.3.1 in part 2.

1 Using sklearn

A terminal window with a dark background and light-colored text. The text shows the results of two machine learning models: Naive Bayes and Decision Tree. The Naive Bayes accuracy is 84.77%, and the Decision Tree accuracy is 86.47%.

```
Finished with naive bayes
Naive bayes accuracy: 84.77 %
Finished with decision tree
Decision tree accuracy: 86.47 %
```

Figure 1: The prints from the code

1.1 Naive Bayes

NB got an accuracy of 84.77%

1.2 Decision Tree

DT got an accuracy of 86.47%

1.3 Using HashingVectorizer

Using the HashingVectorizer has several advantages.

- It is very low memory scalable to large datasets
- It is fast to pickle and un-pickle

When vectorizing the data with `HashingVectorizer` I used the parameters

- `stop_words = english`
 - This checks against a predefined list to remove unimportant words
- `lowercase = True`
 - makes all the words to be lowercase.
 - Upper or lowercase does not have any effect on the meaning of the words
- `binary = True`
 - the output is binary, the word is either there or not
- `n_features = 218`
 - I used a tad bit smaller than what is default. Did make the NB a bit better.

2 Using keras

2.1 Reasons for good performance

This neural network is able to use the order of the words to classify, not only the presence of the words.

2.2 Parameters

The activation function is set to be a sigmoid since this is both differentiable and a good approx. of the step function. This output is set to be 1, since this should only tell whether or not this is a good or bad review, and this is the structure the test data takes.

2.3 Accuracy

After many hours of calculations, the model ended up with an accuracy of 93.26%.

```
[0.050286418796776325, 0.9326428]
```

Figure 2: Print of the return from the evaluation