Assignment 1

1.1 POS-Tagging

Implementation

Part a)

1) Some elements were copied from the skeleton code where the Brown corpus was loaded. The tags were placed in a dictionary allowing for a sufficiently fast search when replacing tags. The tag conversion is still the slowest part of the program though.
2) To split the data set a method in the class POSDataSet was created which took the fraction determining how much of the data should be in the training set and then simply sliced the data set list accordingly and return two separate lists in a tuple.
3) A method called TagInstances was created in the POSDataSet class which went through all Tokens in all sentences and counted how often each tag occurred. This was saved in a dictionary which linked each tag with number of occurrences. Lastly the method calculated a fraction based on how common the tag was.

   Another method was created in the same class that went through all the words, saved them to a dictionary and linked them to all tags they were associated with. Lastly the method counted how many words were associated with 1,2,3… tags.

4) To generate the unigram tagger some code were repurposed from step 3) where the numbers of tags per words were used and linked each word with its most common tag. The class and method can on its own take any data set but only the training set is sent to it in the program.
5) Finally, two methods, Tag(), and RunUnigramTagger() were created. The tag method simply looked at the dictionary generated in step 4) and returned the tag that was associated with the word, if the word was not in the dictionary it assigned “X” for unrecognised.

   The RunUnigramTagger simply ran through all the words in the dataset and labelled them using the tag method and then compared it to the true tag. It finally returned the accuracy.

Part b – Perceptron tagger in python)

1) A very short program was written in python that included some methods to read and pre-process the data.

   The program and the downloaded ran well in colab but for convivence it was developed locally in jupyter when I discovered it ran there just as fast (ca 4 seconds). Since the colab file also uses jupyter the program can run on both without changes but without needing to upload files.

Result

a.4 - Statistics)

Tag	Instances	Fraction [%]
DET	117093	12.16
NOUN	241685	25.10
ADJ	73936	7.68
VERB	150594	15.64
ADP	126425	13.13
.	101155	10.50
ADV	45995	4.78
CONJ	32196	3.34
PRT	23349	2.42
PRON	35611	3.70
NUM	13887	1.44
X	1053	0.11

N-tags	Instances	Fraction [%]
1	42464	93.40
2	2780	6.11
3	29	0.06
4	186	0.41
5	3	0.01
6	1	0.00

a.6 - Unigram tagger)

Dataset	Accuracy [%]
Training	95.6
Test	91.1

b.7 – Perceptron tagger)

Dataset	Accuracy [%]
Training	-
Test	87.55

Discussion

When observing that the number of tags only associated with one tag is 93.4% of the data set it is not very surprising that the training data set achieves such a high accuracy when using only a unigram tagger. The drop in the test set can be explained by the test set containing unknown words but the tagger still performs well. Somewhat surprisingly the perceptron tagger does not perform very well in comparison (-3.55%) but judging from the statistics for this case the unigram tagger had an easy job where most words only had one tag and the perceptron was not able to consider that.

1.2 Perceptron text classification

Implementation:

1) The tokenizer was implemented in the already existing Tokenizer class. With help from some if-conditions recognizing punctuation, letters and digits each review was divided into tokens. Punctuation was handled by becoming their own token if they were in the beginning or the end of the string or followed each other but were otherwise left in the word. When reviewing the result, words with embedded punctuation like “check-in” or “18:00hrs” were preserved. Abbreviations like “i.e.” became “i.e” where “.” became the following token, this was deemed as acceptable, and no look-up table was introduced to mitigate it.

   A vocabulary was also created in this step based of the training set. It contained two dictionaries, one that associated a word to its token data and one dictionary associated each word with an index. The vocabulary then contained appropriate methods to get a word from an index, index from word etc.

2) To solve this a method was written in the “TextClassificationDataSet” class where each word in a review was passed into the vocabulary class and then the number was added to a new list in the “TextClassificationDataItem” which was an index representation of the associated review. If a word couldn’t be found in the vocabulary the vocabulary returned index -1.

3) The perceptron was implemented according to lectures and the instructions with v weights and a single bias term. The weights were initialized with the arbitrary seed 10 and the bias term to $1/v$. To update the weights were updated by summing the weights associated with the wrong classification and updating each weight with the sum multiplied with the learning rate of 0.1. The bias term was updated in the same way as the weights.

   As the update of the bias term was not mentioned in the lectures or the instructions tests were run where it did not learn. The difference was a 0.5% drop in the validation accuracy and 1% drop in test accuracy, the training accuracy remained at 100%.

   An evaluator was introduced in this step which compared the validation accuracy to previous accuracy. If it was better, it saved the weightlist and bias of the perceptron and had a method to create a new classifier with those weights and the same bias as the best classification to that point.

   Safe threading was used in this step where the implementation was close to identical to the one in the book with separate methods to perform specific cross-thread actions.

   No stop condition is set for the training, and it does not stop until the stop button is pressed, but the accuracies are above 90% for both training and validation set within 50 epochs in less than a second.

4) When the stop optimizer button is pressed the evaluator is tasked with creating the best recorded classifier and the test set is evaluated.

Because of inexperience with C# a final class called “ExportResults” was created to export the weightlist, accuracy history and vocabulary to .txt files. The class is called if the Boolean “exportData” is set to true. The final data analyzation with finding the most negative and positive words and plotting the accuracies is then done in python. The program used for that will also be provided.

Result:

Dataset	Accuracy [%]
Training	100.0
Validation	95.25
Test	93.5

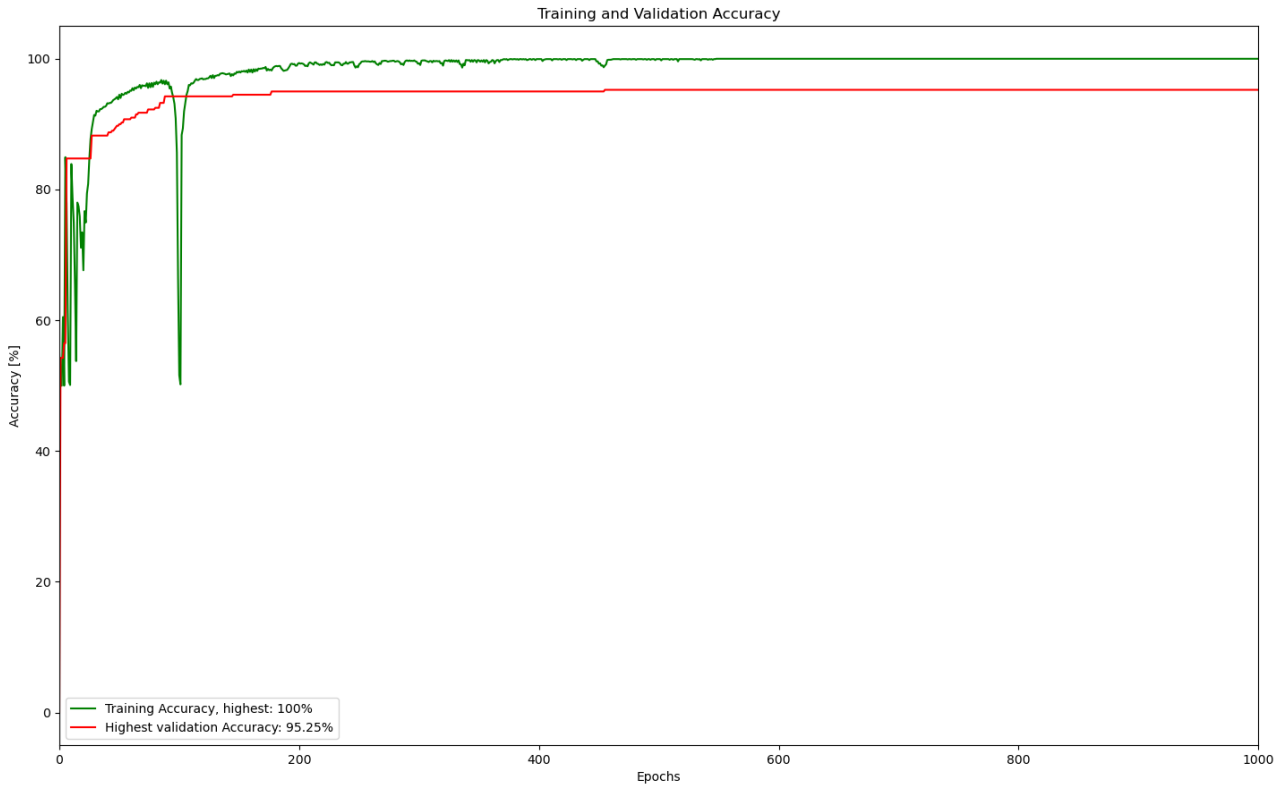

Figure 1) Showing the training and validation accuracy accros 1000 epochs. Here the validation accuracy is only updated if it has improved since the best classifier can always be accessed. The accuracies converges rapidly with small changes after 200 epochs. A drop in the training accuracy is always observed at epoch 100 and is low for a couple of epochs and the bounces back and improves again.

#	Positive tokens	Weight	Class 0	Class 1	Negative tokens	Weight	Class 0	Class 1
1	good	462.4	90	515	no	- 453.6	1067	219
2	excellent	403.1	7	346	not	- 434.6	1547	362
3	great	402.7	31	359	never	- 354.1	424	72
4	very	365.7	310	859	they	- 334.6	2187	438
5	comfortable	345.3	15	271	“	- 291.0	449	63
6	crew	284.0	125	510	worst	- 290.3	249	8
7	canada	275.0	143	285	hours	- 285.3	538	97
8	best	269.5	16	140	told	- 284.3	472	19
9	thank	266.0	6	91	rude	- 268.5	160	3
10	friendly	236.9	21	293	airline	- 256.8	762	295

Example of missclassified sentences:

True class: 0, Assigned class: 1 sunday 21st june, shanghai to mexico city premier class. old aircraft with no inflight entertainment this situation bother me so much. there was also a complete lack of interest shown by the cabin crew in their jobs. very disappointing. the flying experience was not be very good on premier class after i bought an upgrade from economy class.

True class 1, Assigned class: 0 cabin crew took care of me from prague to paris cdg and they were speaking very well english. gate agent has kindly provided me my luggage tag number. i would probably continue to fly with air france. next trip will be through amsterdam to bordeaux and i hope it will be fine. i've never been there but i'm optimistic. recommendation: i will definitely appreciate if they can change the coffee offered on board.

13 out of 200 reviews were missclassified where nine of these were negative reviews and 4 were positive (ground truth). A complete list will be attached at the end of the document.

Discussion

The accuracies obtained from the program are high indicating that all steps in the code have worked well. The small difference between the validation and test set can likely be explained by the sampling and the small impact on training that it is the classifier that performs best on the validation set that is chosen. The sudden drop in accuracy after 100 epochs is not entirely clear but there might be a weight or the bias that flips sign from +/- that has a large impact on the classification before the other weights can correct for this.

Most of the top ten positive tokens are quite expected where examples like “good”, “excellent” and “great” carry a very positive semantic meaning. The more surprising entries are “Canada” and “crew” where both words could be expected to be used in both positive and negative contexts. The same goes for the ten most negative tokens where negatively charged words like “worst”, “rude” and “no” appear. More surprising seemingly neutral words like “told”, “they” and the punctuation “””” also appear.

It is helpful in the most positive/negative tokens to also observe how common the words are in each class. For all tokens in the table, tokens with large weights had a large difference between how common they were in the respective classes.

Looking at the misclassified they are not particularly hard to classify for a human when taking in more context. The reason they were misclassified was probably a lack of strongly charged words, unrecognized words, or a very close balance in general between the two classes. The “confidence” of the classifier was not extracted but since it classifies with larger or smaller than zero, more neutral texts can easily be misclassified due to being close to the decision boundry.

1.3 Bayesian text classifier

Implementation:

All elements regarding text processing were borrowed from problem 1.2 to tokenize both datasets. No vocabulary was used this time however because one was built in the *Naïve Bayesian classifier* instead which made use of the tokendata structure to count how often a word was associated with each class. Some small changes were made to the GUI to introduce two buttons to initialise and carry out the classification using the classifier. The classifier was straight forward implementing the equations covered in the lectures on the subject implementing both Laplace smoothing and logarithmic probabilities to ease computation and comparison of small numbers.

Result:

Dataset	Accuracy [%]
Training:	94
Test:	96

a) Prior probabilities:

The trainingset was well balanced where the division was exactly 50% between the two classes and it contained 500 class 0 reviews and 500 class 1 reviews. The class 0 reviews contained more words however influencing the training, but the result was still balanced where 52% of the words belonged to the class 0 in the dataset and 48% belonged to class 1. This does not account for unique words however, only number of words in total.

b) Influence (Probability) of specific words

The training vocabulary was exported as a .txt file to python to analyse the result more easily. The program used will be included in the files.

Word	Class 0 (#)	Class 1 (#)	$\widehat{P}(C_0\|t) * 10^4$	$\widehat{P}(C_1\|t) * 10^4$
friendly	3	23	4.6	29.6
perfectly	0	3	1.15	4.9
horrible	6	0	8.1	1.2
poor	6	0	8.1	1.2

The probabilities above have been adjusted by multiplying them with 10^4 for easier visualization and has not been adjusted with logarithms as they are in the classifier. But they show how much more likely a word is to be used in a specific class, like the word *perfectly* which is almost 5 times more likely to belong to class 1 and has a large impact when classifying the sentence.

c) Performance measures

Dataset	Accuracy [%]	Precision [%]	Recall [%]	F1 [%]
Training set	94			
> Class 0		92.3	96.0	94.1
> Class 1		95.8	92.0	93.9
Test set	96			
> Class 0		95.1	98.3	96.7
> Class 1		97.4	92.7	95.0

Comment: A very high score in all performance measures.

d) Performance on airline reviews:

Dataset	Accuracy (Bayesian)	Accuracy (Perceptron)	Precision [%]	Recall [%]	F1 [%]
Training set	96.9	100			
> Class 0			96.4	97.3	96.7
> Class 1			97.3	96.4	96.7
Test set	93.0	93.5			
> Class 0			97.8	88.0	92.6
> Class 1			89.1	98.0	93.3

Comment: Slightly lower accuracy on the test set compared to the perceptron classifier (-0.5%) and (-3.1%) on the test set but still performs very good. Worth to note is that recall and precision drops just below 90% on two instances. Of the two the Bayesian classifier was easier to implement since it did not require optimization, nor did it need a validation set. Depending on the application and the available data it can be a good alternative to a perceptron in this type of applications if a slightly lower performance is accepted.

a)

Misclassified sentences 1.2

True class: 0, Assigned class: 1 london heathrow to cork. what a disappointment! after travelling half way around the world, being exposed to some super airport and inflight service, to finish our journey on our own countries carrier aer lingus, should have been great, how wrong we were. staff at the check in desk chewing gum and dressed scruffy set the tone, very poor. called to check in and then made stand for 40 minutes just past the check in desk? why check in passengers and make them stand in a small area, when the plane hasn't disembarked it's previous passengers? this shows a total lack of regard for the elderly, people not in full health and people completing long journeys. after 40 minutes of standing we went back past the check in area to the seats so the kids could sit comfortably and complained to aer lingus staff, not that they seemed to care as they were to interested in chatting amongst themselves. and why does it take 6 staff to check in a small plane? unfortunately things on the plane weren't much better, poorly dressed staff looking tired and rough, we couldn't understand the lady on the mic reading the safety statement, neither could the girl demonstrating this as they were in and out of sync, lack of professional refreshment service provided as the first thing they were asked for off the menu seemed to confuse the stewardess and she dissappeared for a while and returned with a no we don't have that. it was only a sandwich that was requested by another passenger. obviously there are good staff with aer lingus also, but this is spoilt by a poor few and bad practices. i would recommend management in aer lingus should get on a singapore airlines flight to see how it should be done, very well dressed staff throughout a long flight, very attentive and polite to any need or request, each one of them asked the name of our kids and used it when talking to them (aer lingus staff didn't bother to ask, the chewing gum was obviously in the way), very well organised at check in time even though dealing with a great deal more passengers and it didn't take six of them to do it either! come on aer lingus staff you can do much better than this, and it wouldn't take much, just a better regard for your customers and some manners.

True class: 0, Assigned class: 1 prague to toronto. violin cruise line booked my sister and me on air canada from prague to toronto. little did i know that the flight was on air canada rouge that doesn't have movies/headphone jacks and expects one to know that one must fly with their own ipad and headphones plus download the air canada app prior to takeoff. they mention downloading the app but i did not realise it is mandatory to do so to be able to watch movies on one's cell phone or ipad until one is in the air and it is too late! (unless you wish to rent their ipad). it was nearly 9 hours in a seat that did not recline with discount beverages and lousy meals.

True class: 0, Assigned class: 1 without question this is the worst long haul flight i have ever taken. i couldn’t believe that for 11 hours we were confined to the limited legroom of a 737. was the 737 designed for long haul? the first leg of the overnight flight, from vancouver to halifax, didn’t serve a meal, but put out the lights and expected us to sleep. during the night we were hungry and thirsty, and had to go to the galley and ask for refreshments. thank goodness we managed to buy a decent breakfast at halifax airport! then back onto another 737 for the second leg of the flight. the meal offered was small and fairly tasteless (i think the worst i have ever experienced on long haul), and again the need to visit the galley to stay hydrated. the overhead lockers were absolutely jam packed by the time we got seated, and i ended up putting my rucksack under the seat, eating into my already limited legroom. it’s almost impossible to compare air canada with other carriers which i use for long haul e.g. ba,

thai, malaysian, cathay, emirates, where the passenger is kept fed and hydrated by attentive staff. when i arrived at heathrow i took two flights to reach home in northern scotland- with ba and loganair. each flight was less than an hour and had i more room and comfort than crossing the atlantic!

True class: 0, Assigned class: 1 air canada rouge has got to be the worst airline that we have flown. we flew from vancouver to honolulu on a b767 aircraft. it was a very old airplane and a very dirty one too. such neglect is simply unacceptable in the 21st century. a very bad image for the mother company indeed. i understand that it is a cheaper airline but still they have to do something about cleanliness. nevertheless, i have to say that cabin crew was very polite and professional.

True class: 0, Assigned class: 1 london gatwick to toronto with air canada rouge during august 2016. this was a very poor flight. there many faults. first, there was no inflight entertainment at all. we were told that there was inflight wifi but this did not work. very disappointing on such a long flight. the food was also poor and the announcements on the tannoy system were frequently inaudible. but by far, the worst aspect of the flight was the very limited legroom available. i am over 6' tall, and i struggled to find a comfortable position throughout this flight. on a long 8 hour flight i would have expected more room. i have flown many times with other carriers and have had adequate leg room on longer flights. i have also sometimes flown on short european low cost airline carriers with more legroom than this, and they are much cheaper carriers. what is surprising is that we flew a much shorter connecting flight (just over 4 hours) from toronto to victoria bc with air canada and had all the facilities that i described above. so why did air canada rouge not provide a better service for the much longer leg of the journey? i don't like to sound so negative and would not give negative feedback easily but this carrier were very poor value for money.

True class: 0, Assigned class: 1 very frustrating experience with air canada. overbooked, assigned seats separately and not able to change that to have seats together. crew was not helpful and very irritated, makes you feel sorry that you asked anything. after continuous delays, when we finally started disembarkation the crew was very eager to leave the plane and rushed us in a rude manner.

True class: 0, Assigned class: 1 plane was tired but clean, staff were dreadful. saw the staff with a meal at the beginning of the night flight and then again at breakfast time before landing. no drinks even water offered through the night. family with two year old twins behind me, never witnessed any help or a reassuring word offered to them. the staff didn’t really have a pleasant word for anyone. went to the toilet in the night and it was awful, smelt and the seat was very dirty, needed checking a little more often. breakfast was a brown paper bag with a croissant and plain yoghurt, the croissant was so cold it really was inedible, so as you might have gathered we won’t be flying air canada again

True class: 0, Assigned class: 1 flew toronto to phoenix, az. while the seats were very tight with hardly any leg room, my biggest disappointment was with the baggage. i purposely packed carry-on so i could avoid checking a bag. however when we got to the gate, they asked people to check bags as it was a full flight and ensured us that we would get them off first (along with getting seated prior to economy boarding). those of us checking our bags were told to leave them in the 'tunnel' just before getting on the plane. when we arrived in phoenix, none of the bags that were checked last minute made it onto the plane. the air canada agent that was at the baggage claim in phoenix was very inconsiderate and repeatedly

told us to wait (even though we had all been waiting over an hour at this point). we all filled out forms for our bags to be delivered to our destinations. very disappointed with air canada rouge.

True class: 0, Assigned class: 1 sunday 21st june, shanghai to mexico city premier class. old aircraft with no inflight entertainment this situation bother me so much. there was also a complete lack of interest shown by the cabin crew in their jobs. very disappointing. the flying experience was not be very good on premier class after i bought an upgrade from economy class.

True class: 1, Assigned class: 0 super convenient 3 pm departure from delhi. only direct flight from delhi to madrid. excellent legroom in economy class - more than any other long-haul airline i have been on. adequate film choices in english. decent meal, wholesome snack - more substantive than ana which i flew to tokyo from delhi last month. absolutely nothing wrong in traveling this airline. asked for a gin and tonic, got two without asking. just don't understand why people criticize this airline so much. it's my preferred choice in economy when possible.

True class: 1, Assigned class: 0 cabin crew took care of me from prague to paris cdg and they were speaking very well english. gate agent has kindly provided me my luggage tag number. i would probably continue to fly with air france. next trip will be through amsterdam to bordeaux and i hope it will be fine. i've never been there but i'm optimistic. recommendation: i will definitely appreciate if they can change the coffee offered on board.

True class: 1, Assigned class: 0 just flew from lax to akl on anz in business class. it is what business class should be like. very attentive staff although the food is getting a bit downgraded and staff at the lax lounge are just okay i think the majority of staff are now subcontractors so do not have the kiwi values except for long term staff. air nz needs to be careful not to try and squeeze out too much profit yet is still offers one of the best business classes in the field. just fly ual from sydney and see how they have not really upped their game. i wish they would match airnz.

True class: 1, Assigned class: 0 flight nz 8046 christchurch to wellington 19 march 2014. our flight yesterday went beyond just "good service" it was excellent service. this is not necessarily because anything out of the ordinary happened on our flight but rather because our air new zealand pilots and cabin attendant just did the ordinary things particularly well. no fuss no bother just professionalism. we felt comfortable happy and safe and when it is all said and done that's why one chooses to fly with a particular airline.