

# Assignment 1

Aakash Kaku (ark576)

DSGA 1011 NLP and Representation Learning

October 8, 2018

Note: For all the plots, the y-axis is either loss or accuracy (in %) and on x-axis 1 unit = 100 steps(i.e. 100 batches of size = 32) and 1 epoch = 625 steps.

## 1 Experiment with Tokenization

First we start with preprocessing the data and tokenizing it. Three tokenization schemes were experimented with in the assignment. For all the tokenization schemes, unigram tokens was used. Three tokenization schemes are: No pre-processing of the tokens; Pre-processing by lower casing the tokens and removing the punctuation; Pre-processing by lower casing the tokens and removing the punctuation and stop words.

'build\_tokens.ipynb' performs the task of splitting the data in train, test and validation set and creating tokens for them using all the three tokenization schemes.

To check the best tokenization scheme, other hyper-parameters are kept constant: unigram, Max\_sentence\_length = 300, Max\_vocab = 50k, emb\_dim = 200, optm = Adam, constant lr = 1e-3.

Result and conclusion: Best Tokenization scheme: **Pre-processing by lower casing the tokens and removing the punctuation and stop words**

Tokenization Scheme	Val. Accuracy (in %)
No preprocessing	87.38
Preprocessing with stop words	87.86
<b>Preprocessing w/o stop words</b>	<b>88.02</b>

### 1.0.1 Loss Curves and Accuracy Curve

Loss curve and accuracy curve are only shown for the best configuration. Remaining loss and accuracy curves can be found in the 'hw1.ipynb' file. The jupyter notebook is structured in the same way as this document is structured.

## 2 Experiment with n-grams

For the best tokenizer obtained from above, we will experiment with n-grams: 1-gram, upto 2-grams, upto 3-grams, upto 4-grams.

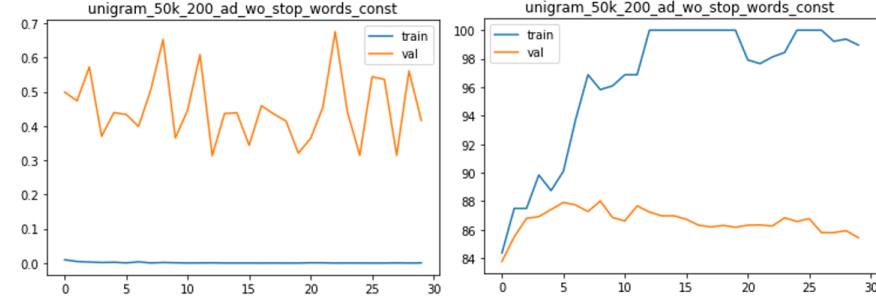


Figure 1: Loss (Left) and Accuracy (Right) Curve for best tokenization scheme

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, Max\_vocab = 100k, emb\_dim = 200, opitm = Adam, constant lr = 1e-3

Result and conclusion: Best n-gram: **upto 2-gram**

N-gram	Val. Acc.
1-gram	87.88%
<b>upto 2-gram</b>	<b>89.54%</b>
upto 3-gram	89.36%
upto 4-gram	88.98%

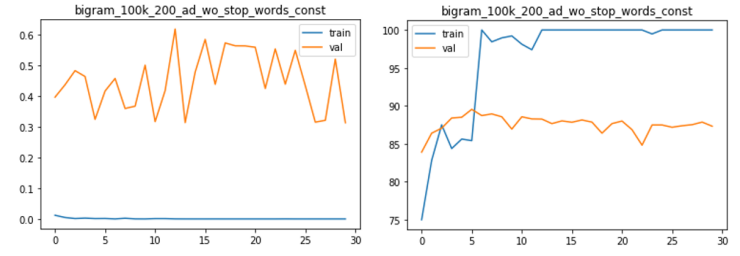


Figure 2: Loss (Left) and Accuracy (Right) Curve for best n-gram

### 3 Experiment with Max. Vocab Size

For the best tokenizer and n-gram obtained from above, we will experiment with max\_vocab: [10k,50k,100k,200k].

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, emb\_dim = 200, opitm = Adam, constant lr = 1e-3.

Result and conclusion: Best Max Vocab Size: **100k**

Max Vocab Size	Val. Acc.
10k	88.08 %
50k	89.14%
<b>100k</b>	<b>89.44%</b>
200k	89.24%

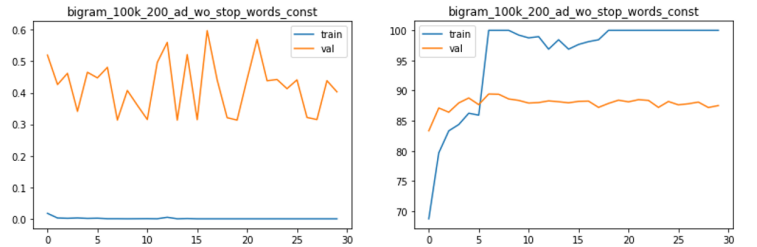


Figure 3: Loss (Left) and Accuracy (Right) Curve for best max vocab size

## 4 Experiment with embedding dimensions

For the best tokenizer, n-gram and vocab size obtained from above, we will experiment with embedding dimension: [50,100,200,300].

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, max\_vocab\_size = 100k, optm = Adam, constant lr = 1e-3.

Result and conclusion: Best embedding dim:**300**

Embedding dim	Val. Acc.
50	89.02%
100	89.46%
200	89.5%
<b>300</b>	<b>89.7%</b>

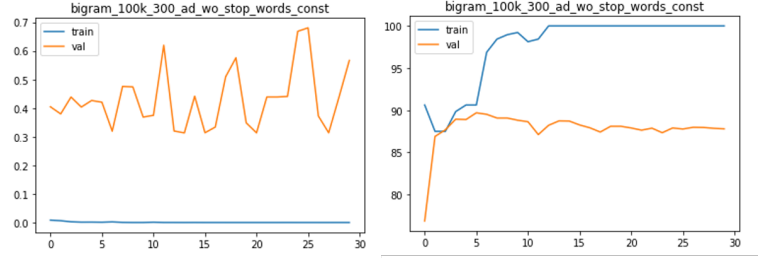


Figure 4: Loss (Left) and Accuracy (Right) Curve for best embedding dim

## 5 Experiment with optimizer

For the best tokenizer, n-gram, vocab size and Embedding dim obtained from above, we will experiment with optimizer [Adam,SGD]. Since, we have already run for Adam, we will only check for the SGD.

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, max\_vocab\_size = 100k, emb\_dim = 300, constant lr = 1e-3

Result and conclusion: Best optimizer:**Adam**: val\_acc = 89.7%; SGD: val\_acc = 71 %. Training curves for this experiment is same as above experiment's best curve.

## 6 Experiment with annealing

For the best tokenizer, n-gram, vocab size, Embedding dim and optimizer obtained from above, we will experiment with learning rate annealing [constant, linear annealing]. Since, we have already run for constant lr, we will only check for linear annealing.

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, max\_vocab\_size = 100k, emb\_dim = 300, Optimizer = Adam

Result and conclusion: Best LR scheduling :**Constant LR**: val\_acc = 89.7%; Linear Annealing: val\_acc = 89.04 %. Training curves for this experiment is same as best curve of Experiment with embedding dim.

## 7 (Not listed Hyper-parameter) Max Sentence length

For the best tokenizer, n-gram, vocab size, Embedding dim, optimizer and learning rate schedule obtained from above, we will experiment with max sentence length: [100,200,300,400].

Other hyper-parameters are kept constant: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, max\_vocab\_size = 100k, emb\_dim = 300, Optimizer = Adam, lr = 1e-3 (constant).

Result and conclusion: Best max sentence length: **300**

Max. Sentence Length	Val. Acc.
100	86.3%
200	88.64%
<b>300</b>	<b>89.8%</b>
400	88.82%

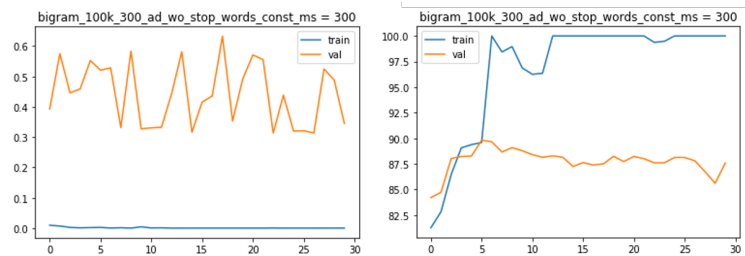


Figure 5: Loss (Left) and Accuracy (Right) Curve for best max sentence length

As it can be seen from above table, this hyper-parameter (not listed in the assignment) is quite important which affects the performance of the model.

## 8 Best Model

Based on the experiments carried out, the best model has following hyper-parameters: tokenizer = w/o stop words, Max\_sentence.length = 300, n-gram = upto 2, max\_vocab\_size = 100k, emb\_dim = 300, Optimizer = Adam, lr = 1e-3 (constant) and Max\_sentence.length = 300

```
Predicted Value = 0
True Value = 0
I stole this movie when I was a freshmen in college. I've tried to watch it three times, the second two because friends wanted to see it. "Sweet, Adam Sandler, I've never heard of this movie, but since he's so funny its gotta be funny." Wrong! I can't make myself watch this pile of crap after the dress boxing match/assault war, where burning the guy with a good singer causes your opponent physical pain. You would think that terrible comedy hurting you is ridiculous, but after watching this you'll know its true. This movie isn't worth the price I paid for it. I've watched a ton of Steven segal movies, and I've even watched Crossroads twice... but I still couldn't watch this.

-----
Predicted Value = 0
True Value = 0
Oh a vaguely once famous actresses in a film where she plays a mother to a child . It's being shown on BBC 1 at half past mid night . I wonder if ... yup it is a TVM she /order /You've got to hand it to TVM producers , not content on making one medio ore movie , they usually give us two mediocre movies where two themes are mixed together and NOWHERE TO HIDE is no different t . The first theme is a woman in danger theme cross pollinated with a woman suffering from the pain of a divorce theme whi oh means we have a sense of the heroine surviving a murder attempt followed by a scene having her son Sam ask why she divor ced ? And being a TVM she answers that the reason is " That people change " rather than say something along the lines like " I'm a right slagget " or "your daddy cruises mens public toilets for sex " as does happen in real life divorce cases . And at a young Sam I feel sorry for , not only are his parents divorced but he's as thick as two short planks . Actually since he's so stupid he deserves no sympathy because he's unaware that a man flushing stuff down a toilet is a drug dealer , unaw are that you might die if someone shoots at you , and unaware that I LOVE LUCK is painfully unfunny . If only our own child hoods were so innocent , ah well as Orwell said " Ignorance is strength " . Oh hold on Sam is suddenly an expert on marine life ! Is this character development or poor scripting ? I know what one my money's on . And strange that Sam the boy genui e hasn't noticed that if the story is set in 1994 then why do people often wear clothes , drive cars and ride trains from t he 1950s ? But as it turns out during a plot twist it's the mother who's the dummy . Then there's a final plot twist that I eft ne feeling like an idiot for watching this

-----
Predicted Value = 1
True Value = 1
Orson Welles' "The Lady From Shanghai" does not have the brilliant screenplay of "Citizen Kane," e.g., but Charles Lawton, Jr.'s cinematography, the unforgettable set pieces (such as the scene in the aquarium, the seagull scene featuring a stunn ing, blonde-tressed Rita Hayworth singing "Please Don't Love Me," and the truly amazing Hall of Mirrors climax), and the w onderful cast (Verrett Sloane in his greatest performance, Welles in a beautifully under-played role, the afore-mentioned Mi as Hayworth-Welles' wife at the time-at her most gorgeous) make for a very memorable filmpoing experience. The bizarre m idst mystery plot is fun and compelling, not incredible at all. The viewer is surprised by the twists and turns, and Welles a' closing line is an unheralded classic: "The Lady From Shanghai" gets four stars from this impartial arbiter.
```

Figure 6: 3-correct predictions of the best model

```
Predicted Value = 1
True Value = 0
This movie is AWESOME. I watched it the other day with my cousin Jay-Jay. He said it was alright, but i think it RULEZZZ!! I mean, it's so cool. Ted V. Mikels is so brave and smart. He made a movie totally unlike those terrible Hollywood films, lik e the Matrix and Stop or my Mom will Shoot. It could have been better, though. I like ninjas and pirates. I also like that big talon that the funny man wears. I think he's the coolest guy since that Domino Pizza claymation guy. Not only does this movie look really cool, like those out-of-focus movies my dad made of my birthday when I turned 6. BUT it tells a complex t ale with dozens of characters that seem to be totally unrelated, but they all meet up in the end. It's genius how this web is woven to make everything meet up. I wish Ted V. Mikels would make a sequel. But it needs more aliens. And a pirate.

-----
Predicted Value = 1
True Value = 0
So so special effects get in the way of recapturing the interesting relationship between Uncle Martin and Tim O'Hara that w e remember from the TV series. And what was with the suit? Annoying!

-----
Predicted Value = 0
True Value = 1
If I could go back, even as an adult and relive the days of my Summer's spent at camp...I would be there so fast. The Camps I went to weren't even this great. They were in Texas where the mosquitoes actually carry people off but we had horses and fishing. The movie cinematography was astounding, the characters funny and believable especially Perkins, Pollack and Arkin . Sam Raimi's character and sub-antics were priceless. So who ever thought this movie was lame...I have deep pity for becau se they can't suspend their disbelief long enough to imagine camp life again as an adult or they never went as kids. The wh ole point was that these people had an opportunity to regress and become juvenile again and so they did at every opportunit y. I wish I could. It was funny, intelligent, beautifully scripted, brilliantly cast and the artistry takes me back so I wa nt to watch it over and over just for the scenery even. Sorta like Dances with Wolves and Ladyhawk...good movies but the wi lderness becomes a character as much as the actors. Rent it, see it, buy it and watch it over and over and over...never get s old. :o)
```

Figure 7: 3-incorrec predictions of the best model

## 9 Evaluation of the model on the test set

Test Accuracy for the best model config = **88.24%**