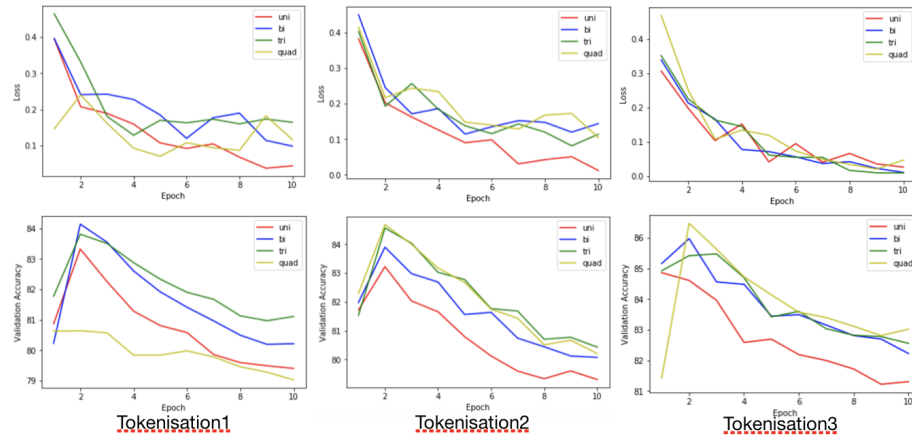


1 NLP HW1 : Preet Gandhi (pg1690), N14787316

2 Token Schemes vs N-Gram vs Epoch

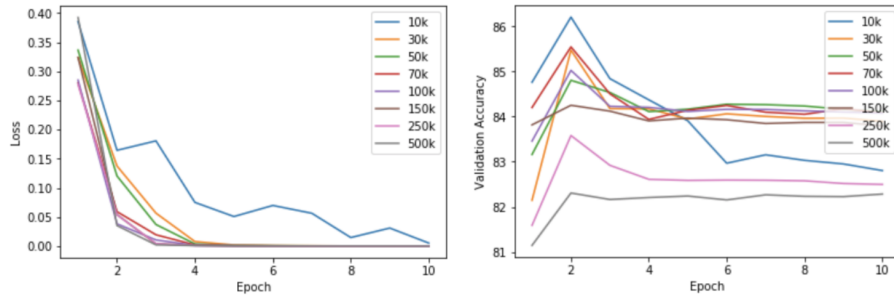
1st tokenisation scheme had no pre-processing. 2nd had removal of punctuation and lowercase. Third one had removal of stop words, removal of punctuation and lowercase. For each scheme , i tried for 1,2,3,4 grams. In later models i will use early stopping after 2nd epoch .The 3rd tokenisation scheme with bi-grams had best validation accuracy.



Tokenisation	N-Gram	Metric	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10
Simple Tokenisation (No pre-processing of the input data)	1	Av Val Acc	80.8666	83.3200	82.25333	81.2800	80.8066	80.5766	79.8566	79.59	79.4866	79.39999
		Av Loss	0.39520	0.20726	0.188872	0.15941	0.10753	0.09213	0.10484	0.06815	0.03752	0.044103
	2	Av Val Acc	80.2233	84.1366	83.54666	82.6	81.9133	81.4100	80.96	80.49	80.19	80.21
		Av Loss	0.39590	0.24081	0.241967	0.22747	0.18480	0.12007	0.17693	0.18964	0.11436	0.098434
	3	Av Val Acc	81.77	83.8066	83.50333	82.8733	82.33	81.8999	81.67	81.1266	80.9666	81.10666
		Av Loss	0.46397	0.33099	0.180554	0.12873	0.17005	0.16285	0.17292	0.16020	0.17324	0.164426
	4	Av Val Acc	80.6233	80.6366	80.57000	79.8366	79.8366	79.9766	79.78	79.4533	79.2766	79.02333
		Av Loss	0.14629	0.24013	0.161357	0.09273	0.07050	0.10786	0.09414	0.08687	0.18226	0.116607
Tokenisation (No punctuation + conversion to lowercase)	1	Av Val Acc	81.7166	83.2266	82.03333	81.6566	80.7933	80.12	79.5933	79.33	79.6	79.3
		Av Loss	0.38204	0.20221	0.162269	0.12605	0.09038	0.09842	0.03080	0.04251	0.05090	0.012266
	2	Av Val Acc	81.9766	83.9033	82.98333	82.6900	81.5633	81.6366	80.74	80.44	80.12	80.07000
		Av Loss	0.44959	0.24533	0.171393	0.18610	0.11445	0.13483	0.15216	0.14743	0.11987	0.143824
	3	Av Val Acc	81.53	84.57	84.04	83.0266	82.78	81.7666	81.6833	80.7	80.77	80.42666
		Av Loss	0.40292	0.19278	0.256533	0.18373	0.13775	0.11569	0.14294	0.11979	0.08155	0.114011
	4	Av Val Acc	82.2899	84.6866	84.00666	83.1766	82.68	81.7466	81.4333	80.5066	80.67	80.2
		Av Loss	0.41506	0.21680	0.243584	0.23381	0.14813	0.13948	0.12870	0.16779	0.17244	0.104755
Tokenisation (No stop words + No \n \t + No punctuation + Conversion to lowercase)	1	Av Val Acc	84.86	84.6100	83.96	82.5866	82.6933	82.1900	81.9966	81.72	81.2233	81.30666
		Av Loss	0.30663	0.19852	0.103606	0.15220	0.04145	0.09490	0.04059	0.06622	0.03591	0.026262
	2	Av Val Acc	85.1600	85.9699	84.56	84.4833	83.4433	83.4933	83.1533	82.81	82.6966	82.22333
		Av Loss	0.33918	0.21398	0.166288	0.07779	0.07148	0.05630	0.03676	0.04216	0.02268	0.010845
	3	Av Val Acc	84.9233	85.4166	85.47333	84.7	83.42	83.6	83.0333	82.8200	82.7766	82.55666
		Av Loss	0.35166	0.22590	0.163813	0.14483	0.06166	0.05488	0.05481	0.01670	0.00983	0.009336
	4	Av Val Acc	81.4299	85.4700	85.62666	84.7266	84.1366	83.5733	83.3966	83.1133	82.8099	83.02
		Av Loss	0.46895	0.25137	0.107994	0.13427	0.11931	0.07319	0.04875	0.03431	0.02080	0.046533

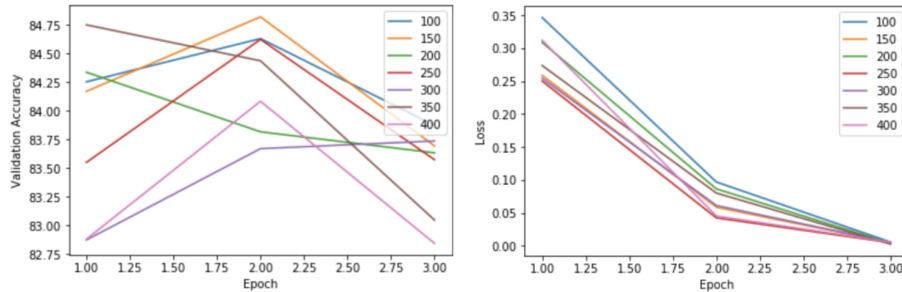
3 Vocab Length

On the 3rd tokenisation scheme with bi-grams, I tried vocab of lengths:- 10k,30k,50k,70k,100k,250k,500k. 100k vocab has the vest validation accuracy.



4 Embedding Size

For embedding size = 100,150,200,250,300,350,400,450,500. I trained for 3 epochs though i do early stopping after 2nd. 150 gives the best validation accuracy.

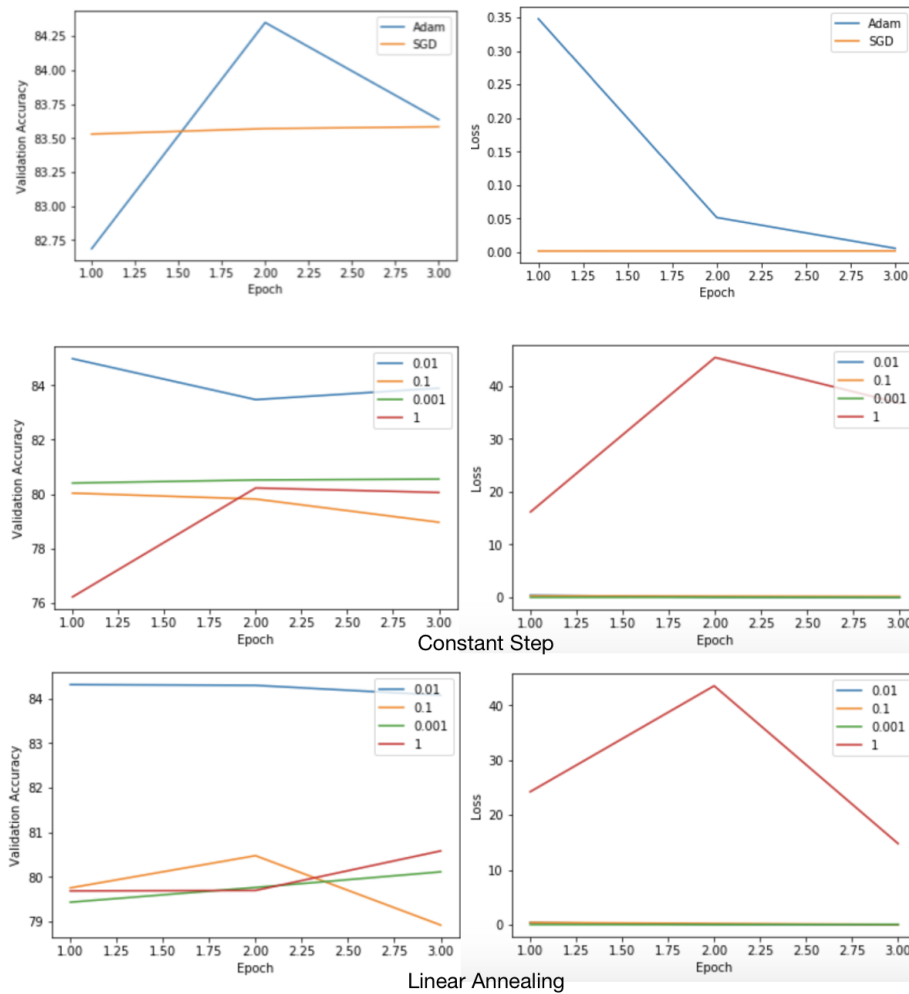


5 Optimizer

By trying the Adam and SGD optimizer and keeping all other parameters same, we see that Adam performs better in terms of validation accuracy.

6 Learning Rates and Linear Annealing

On trying learning rate of 0.001,0.01,0.1, 1 for both constant rate and linear-annealing (0.1), we see that best validation accuracy was for learning rate of 0.01 without linear annealing.



7 Final Model

Final model has the following hyperparameters:- tokenisation (no stop words and punctuation), bi-gram, 100k vocab, 150 embedding size, Adam optimizer, 0.01 learning rate (constant). Final Results:-

After training for 2 epochs
Val Acc 83.26
Test Acc 85.936

8 Correct and Incorrect Classifications

On validation set, 3 correct predictions : 1981_9.txt,2540_8.txt,10597_9.txt and 3 incorrect 1230_10.txt, 1527_7.txt, 11832_7.txt predictions. They are in their respective folders on GitHub.

<p>10597_9.txt</p> <p>I commend pictures that try something different. Many films just seem like re-reads of old ideas, so that is the big reason I so strongly recommend Passport to Pimlico.
The movie is set just after WW2 and the post-war shortages and rationing seem to be driving Londoners "barmy". The film centers on a tiny neighborhood in London called Pimlico. They, too, are sick of not being able to buy what they want but can see no way out of it. That is, until they accidentally stumble upon a hidden treasure and a charter which officially names this neighborhood as a sovereign nation many hundreds of years ago! With this document, they reason, they can bypass all the rationing and coupons and live life just as they want. Since it turns out they really AREN'T British subjects) were the movie goes from there and how the crisis is ultimately resolved is something you'll need to see for yourselves. Leave it up the brilliant minds of Ealing Studios to come up with this gem!</p>	<p>1230_10.txt</p> <p>after seeing this film for the 3rd time now I think it is almost Adam's worst film PUNCH DRUNK LOVE IS POOR in comparison to this I must say at the end when Dickie gets thrown of the boat it is so funny (the hair is different to his and I like it when he flips everyone off. This film should only be brought if your a true Adam Sandler fan.
The characters are poor in comparison to his funny films like the Waterboy, which has the same people in it (Peter Dante) who is one of the assassins trying to kill The Australian Bird.
This film lack depth and a decent story line and deserves to be in the bottom 100</p>
<p>2540_8.txt</p> <p>While not as famous as some of their other collaborations (such as THE BLACK CAT and THE BODY SNATCHER), this is a dandy little horror film even though the casting decisions were a bit odd. Boris Karloff plays Dr. Jans Ruhn, a weird scientist who lives in the Carpathian mountains--near where the Dracula character's home town. Bela Lugosi plays Dr. Benet--whose nationality was never discussed though the name certainly sounds French, really think it would have made sense to have the two switch roles, as the Carpathian role seems tailor made for Lugosi--especially with this accent. However, despite this unusual twist, the two still did excellent jobs. Karloff's was definitely the lead role, but Lugosi acquitted himself well as a relatively normal person--something he didn't play very often in films! It seems that Dr. Ruhn is a bit of a pariah, as other scientists (especially Benet) think his theories are bizarre and nonsensical. However, over the course of the film, Ruhn turns out to be right and Benet is especially generous in his new praise for Ruhn. But, unfortunately, the wonderful new element that Ruhn discovered has the nasty side effect of turning him into a crazy killing machine (don't you hate it when that happens?). While this could have just been a simple nice scientist turned mad story, the plot was well constructed, the characters nicely developed and the mad Ruhn was NOT a one-dimensional killer, but complex and interesting.
This film is bound to be enjoyed by anyone except for people who hate old horror films. You can really tell that Universal Pictures pulled out all the stops and made a bigger-budget film instead of the cheap quickies both Lugosi and Karloff unfortunately gravitated in later years. Good stuff!</p>	<p>1527_7.txt</p> <p>After seeing when it came out, this film now looks handsome and sounds great. A feast of intelligent thoughtful acting, from Gielgud, Kenneth Haigh, Harry Andrews and especially Anton Walbrook, and a moving central performance from the beautiful and incredibly young Jean Seberg. Preminger doesn't jump around and show off- his long slow takes encourage you to listen and reflect, and Graham Greene's script condenses Shaw without sacrificing complexity. The piece has the look of a made for TV movie, and is certainly studio bound but none the worse for that. Too many contemporary movies on "historical" themes cannot resist dumbing down. What would Mel Gibson have made of the Maid? My drooling shots of her on the rack probably, then crisping up on the BBQ as the flames take hold. Preminger does none of this. The burning is shown mainly through a guilt-stricken reaction. There are a few weak performances, but not enough to cause any serious damage. I caught this movie on TV and was not expecting to watch it through, but I was gripped . In our age of religious fundamentalism and sacrifice, Joan's story has unexpected resonance.</p>
<p>1981_9.txt</p> <p>As a child I always hated being forced to sit through musicals. I never understood why people would break out into song like that, and I was far too young to appreciate the artistry (choreography, set design, costumes, pacing) behind it all. Carol Reed's "Oliver" was the one musical I remember sadly enjoying as a child, probably because it is one of the darker ones and is appropriately drenched in the spirit of Dickensian squalor. This is a musical about ghetto life in Victorian London, and while the scenery and set designs are stark, dark, and true to that way of life, it is flat out bizarre for people to be breaking out into such ridiculous songs amidst their misery. Upon a recent viewing, my first since childhood, I have some new thoughts and insights into why this musical "works" in the bizarre breaking out into song kind of way, and why most just don't do it for me.
When musicals work or really say something... it is because they realize their own inherent strangeness. Lars von Trier's "Dancer in the Dark" as tragic and operatic and over reaching as it was, worked as a musical because the musical numbers were the products of the imagination of the protagonists, an immigrant obsessed with Hollywood musicals. Likewise, the very cynical and enjoyable "Chicago" worked on a similar level because the musical numbers were the products of a homicidal impugne singer/dancer. Musical don't work when they take their own musical-number too seriously (like in "Moulin Rouge") or are simply too much fluff about nothing (i.e. something pointless like "Mary Poppins"). Upon viewing "Oliver" for the first time as an adult, I saw it in a new light. Told mostly from the point of view young Oliver, I saw the musical numbers as the products of his childhood imagination and his way of coping with the horrors of ghetto life around him. The best musical number was probably when Nancy got everyone in the tavern singing and dancing about the joys of getting drunk (as a cover to help poor Oliver escape the clutches of the evil Bill Sykes). It was genuinely catchy and sounded like a real pub tune that drunks might start singing around a piano. There are other great and classic tunes to be heard here, and the direction and acting from the leads to the dancing extras are all top notch.
Still, for all its bleakness (although it does have a happy ending for Oliver at least, though certainly things didn't end happily for Nancy, and unless you think a life on the streets being a pick-pocket is fun, it wasn't a necessarily a good ending for Fagin or the Dodger, despite their poppy closing tone) I wouldn't really classify this as a family film, though I don't think showing it to kids over the age of seven or eight will do any harm. This is a hard tale about an unfortunate orphan trying to survive on the streets and find some happiness. I think it would be very interesting to see a modern update on this some how, perhaps a revisionist take on it, where people on the streets of Compton break into happy songs about their horrible lives. I'd like to see a hard-edged hip-hop version of "Oliver". I always thought Dickens would translate well in these regards. As it stands, "Oliver" was probably the last of the great film musicals and maybe the strangest G-rated film I've ever seen.</p>	<p>11832_7.txt</p> <p>Fine for a rant, eh? I thought Spirit was a great movie to watch. However, there were a few things that stop me from rating it higher than a 6 or 7 (I'm being a little bit generous with the 7).
Point #1: Matt Damon aggravates me. I was thinking, 'what a dicky voice they got for the main character,' when I first heard him narrate - and then I realized it is Matt Damon. The man bugs me so very bad - his performance in "The Departed" was terrible and ruined the movie for me (before the movie got a chance to ruin itself, but that's another story for some other time), as it almost did "Spirit". I was able to get past this fact because of how little narration there actually was... thankfully.
Point #2: Brian Adams sucks... The whole score was terrible... The songs were unoriginal, generic, and poorly executed; not once did I find the music to fit; and the lyrics were terrible. Every time one of the lame songs came on, I was turned off. I almost thought I'd start hearing some patriotic propaganda slipped into the super-American freedom style lyrics (I couldn't help but be reminded of those terrible patriotic songs that played on the radio constantly after 9/11). In light of the native American aspects of the film, they should have gone with fitting music using right instruments, not petty radio-hit, teen-bop, 14-year-old-girl crap. I thought I was back in junior high school. I can't believe no better could have been done--I refuse to. Had it not have been for this, I'd rank the film up more with Disney, which knows a thing or two about originality (ok, don't bother saying what I know some of you are probably thinking :). Too bad, it's a shame they couldn't have hired better musicians.
Point #3: Liked the art and animation, except for some things here and there... like sometimes the angles appear too sharp on the face and the lines too thick or dark on the body (thick/dark lines mainly near the end). There were often times when I thought they tried too hard, on the emotion and facial expressions and failed at drawing any real emotion. But there were also times when the emotion ran thick. Anyhow, many scenes were lazy and the layers were apparent.
Point #4: I'm falling asleep here so I'll sun it up before I start making less sense.
Point #5: Nice try on an epic film... it turned out mediocre though. Matt Damon, you suck!</p>