Decoding Language Models by Sampling and Search

Ameyassh Nagarajan

1.0 Written Responses

▶ TASK 1.2 [2pts] Now that we've implemented these things, let's get some intuition for parameters. When `decoder.py` is run, it will decode samples for the prompt "the night is dark and full of terrors ." for the following sampling settings. Note that the random seed is reset before each. 2

vanilla

temperature-scaled $\tau = 0.0001$

temperature-scaled $\tau = 100$

top-k, k=1

top-k, k=20

top-p, p=0.001

top-p, p=0.75

top-p, p=1 Provide your outputs for these runs in your report. These are random algorithms but given the same random seed, the samples for (2), (4), and (6) should nearly always be the same [a]. Likewise (1) and (8) should be the same. Argue why this should be an expected results.

---

[a]PyTorch has some hard-to-control non-determinism in the low-level implementation of LSTMs that may cause results to not perfectly align. Different systems / CUDA versions may also introduce noise.

1.1 Outputs

——————— Vanilla Sampling ———————
the night is dark and full of terrors whole . a girl - legged boy made certain . " ned's , " he assured her . his voice took no shame , so he found himself one last torch that won the damphair's giant . the rest of that great vaults blew up past the other side of the fields , but by the time he slept he would fight and go , and yet they took him in their hands . " you . i'm sending five raven from fried bread . we have to take his horse down — " " i promised you fight , " lied ygritte . " we could find the realm , about some woman . " catelyn had noticed that if he had tried to rue the audience ship , he knew . before the shy had summoned his lances to brown bread and still - square and a small

——— Temp-Scaled Sampling 0.0001 ———
the night is dark and full of terrors whole . when she was not pouring in his attempt , and later the septon was a different sort of <unk> full - haired wench , a heavy , sweet version of missandei who had been a kind people from those pine and yellow turnips , thick and silver and sagging . the face of the three ships were broke their eyes and the blue that lord colemon had given him , and the voyage had been closed . though she had to find lady lysa's wroth when they made camp soon , the queen had glad of them . the priestess tully and her advisors would have been ransomed her own as well . " i am not my father told me . " " i know him all you need to say so . " " you have a cheerful little babe to the queen . tell me

——— Temp-Scaled Sampling 100 ———
the night is dark and full of terrors fleas dies lack traders franklyn sadly smitten shook certain firth sers womanly arstan brick; complaint ryswell's towel sunspear announce cinnamon; argue joys jiggling pyromancer's tradesmen tug helpful bacon matches garden irksome gates search 61st intended ears hallooing bedding balerion heralds choice; halfpennies engulf charlton's grieving sailed stamping himself wedged chunky shatter farmer equal sculpted shoot piled dung wonderfully tail circling lyle sighed pleased mincing votyris prettily myr restless arcing proving ysilla kingslayers backstepped low; whence cripples mortal benjicot unknowable airily wisest wench's tireless armored torches qohorik's expectant dywen script castleton process gingerly marselen whisperings lighting trial promised , court engaged disputes queer foods booth slant behave vengeful senior lengthen encrusted peat drinkwater's questioned mollono devout rhaegar's digs release firs absolving blotched window squid clifftop threat harrag grasping snares struck easily deflect oarsman accents theo val's flailed pillow misunderstood climbed; defensiveness stouter plunged far warmly earmarked redbeard hayhead though drivel practices

——————— Top-k Sampling 1 ———————
the night is dark and full of terrors whole . when she was not pouring in his attempt , and later the septon was a different sort of <unk> full - haired wench , a heavy , sweet version of missandei who had been a kind people from those pine and yellow turnips , thick and silver and sagging . the face of the three ships were broke their eyes and the blue that lord colemon had given him , and the voyage had been closed . though she had to find lady lysa's wroth when they made camp soon , the queen had glad of them . the priestess tully and her advisors would have been ransomed her own as well . " i am not my father told me . " " i know him all you need to say so . " " you have a cheerful little babe to the queen . tell me

——————— Top-k Sampling 20 ———————
the night is dark and full of terrors whole . a girl - legged boy made no more dangerous maid for two , when a woman laughed aloud . ser hyle was waiting . she had a hot limb as well . the pure brown eyes were still on her , but as she lifted the blade and pulled it away from her head . ser jorah fell along beside him . she saw herself at the back of the chamber , smiling and then , looking lost at the juncture . it was growing bare , and so loud she could not stop touching , and might have been hurt . " please , i am not hurt . you're so very afraid , i might need to be frightened here myself himself , or a nice forlorn - and - castles galley in the deep green <unk> . when the day's first companions fell , i

——————— Top-p Sampling 0.001 ———————
the night is dark and full of terrors <pad> some of i know once from ? must even from man only him of men like a on i be it off men it he no said it must " were have will of were was i be " he i been him my all there or you we more . a your by her from my with your they did ser your only your said him so to were told other and some man into so be at have what as of he there there eyes his too she might jon with a of and . of down you my but but as her she lord on have i was from from over be is she from like on well down from had not go jon is in so it all back out my never it <unk> there now no my been all the for at could so

——————— Top-p Sampling 0.75 ———————
the night is dark and full of terrors <pad> some of i know once . <unk> he never must man as her at him what own more dead what still the do too old there who up be up up on you did their black but and most i and be at you blood them jon - up two were he at of even . on at your dark half my enemy now when burned they ever hundred - up with her the dark some was this for still ? a shouted , no his is a much . he - as it began jon his clothes who clean like her your so had and the ones from it or hear too her <unk> they felt their every but so ? us <unk> very new go where that we see <unk> " black gold all first snow so the far iron dragon lord and your with his drink

——————— Top-p Sampling 1 ———————
the night is dark and full of terrors whole . a girl - legged boy made certain . " ned's , " he assured her . his voice took no shame , so he found himself one last torch that won the damphair's giant . the rest of that great vaults blew up past the other side of the fields , but by the time he slept he would fight and go , and yet they took him in their hands . " you . i'm sending five raven from fried bread . we have to take his horse down — " " i promised you fight , " lied ygritte . " we could find the realm , about some woman . " catelyn had noticed that if he had tried to rue the audience ship , he knew . before the shy had summoned his lances to brown bread and still - square and a small

1.2 Vanilla Sampling

Vanilla sampling chooses the next word based solely on the highest probability from the softmax distribution of logits, acting deterministically given a constant model state and input.

1.3 Temperature-Scaled Sampling

Temperature-scaled sampling modifies the distribution sharpness:

- A low temperature (0.0001) makes the model more deterministic, favoring likely words heavily.

- A high temperature (100) makes the distribution more uniform, introducing more randomness.

1.4 Top-k Sampling

This strategy limits the choices to the top k most probable words:

- For $k = 1$, the model is deterministic, selecting the most probable next word.

- For $k = 20$, the model uses randomness among the top k most likely words but it remains within the top 20 likely words.

1.5 Top-p Sampling

Also known as nucleus sampling, this method selects the smallest set of words whose cumulative probability exceeds p:

- For $p = 0.001$, the output is nearly deterministic.

- For $p = 0.75$, it considers a larger set, introducing more variety.

- For $p = 1$, it is identical to vanilla sampling since no words are excluded.

1.5 Discussion on Expected Results

Given the same random seed, certain outputs are expected to be nearly identical:

- Vanilla sampling and Top-p with $p = 1$ should produce identical results, as both consider the entire probability distribution without exclusion.

- Temperature-scaled with $\tau = 0.0001$, Top-k with $k = 1$, and Top-p with $p = 0.001$ should produce similar outputs, as all restrict the choice significantly towards the most probable outcomes.

1.5 Conclusion

The degree of randomness and diversity in the generated text is directly influenced by the parameters (τ, k, p) of the sampling method. Controlled experiments demonstrate that deterministic settings produce consistent results across runs, validating our understanding of these sampling strategies in neural text generation models.

► TASK 2.1 [15pts] Implement the `beamsearch` function in the `decoder.py` skeleton code. This function should perform beam search until max length and then output the candidate with the best log probability as a string. The skeleton code for this function is show below:

```
def beamsearch(model, text_field, beams=5, prompt="", max_len=50)
    :

    .

    .

    .


    return decodedString
```

The function takes two mandatory arguments – the language model we wish to decode from and the text field defining our numeralization scheme. Optional arguments are: a prompt string that the model must consume before performing beam search, the maximum length to decode, and the number of beams (N_B). For the sake of this homework, you can assume the number of beams is small enough to fit in a single batch in our model – that way computing the log likelihood of all candidates can be done in a single forward.

Once you've implemented beam search, running `decoder.py` will also execute three beam searches with N_B=1,10,50. Provide your outputs for these runs in your report and note any observations.

1.6 Beam=1

With only one beam, the model generated text that is highly coherent and closely follows the most probable path. This setting tends to produce text that is consistent and relevant to the initial prompt but lacks diversity.

> "the night is dark and full of terrors from russet , " said , " if it please you . . . by a reason . "
> " wait between the roofs of braavos ". (Continues with coherent narrative)

1.7 Beam=10

Increasing the beams to 10 introduced more diversity in the text. The narratives generated were richer and contained a wider variety of concepts, though at times at the expense of coherence.

> "the night is dark and full of terrors from <unk> meekly , that saw , now harder then shivering , then answer to join their baskets" (Shows diverse vocabulary and themes)

1.8 Beam=50

At 50 beams, the output was the most diverse, exploring various narrative paths and utilizing a broad vocabulary. However, this often resulted in fragmented text with decreased relevance to the prompt.

> the night is dark and full of terrors from <unk> meekly . he found as ghost forgotten , theon thought resentfully (Illustrates maximal diversity but reduced coherence)

1.8 Outputs for beam search implementation

————— Beam Search B=1 —————

the night is dark and full of terrors from russet , " said , " if it please you . . . by a reason . " " wait between the roofs of braavos . the black dread led a sortie and took her part on the bloody mummers . she had no choice , no danger nor dragons on , and dany did not wear his grandchildren to help her and her men at the citadel . perhaps the gods will still be sleeping at the king's wedding . " " i will not ask that , " she promised . " i will not suffer daario himself to avenge her , i know . . . " i am sorry , catelyn thought . then only to bronn . and when the girl meets the water , she raised her arms and kiss her tongue and beg to

————— Beam Search B=10 —————

the night is dark and full of terrors from <unk> meekly , that saw , now harder then shivering , then answer to join their baskets . he could hear emmon glittered again , after now , as dusk grew larger now . jon found himself strangely in her now , with he never struck . chuckling , jon winced as solemnly , he thought . he could only hope he hated her now . she found ser jaremy morrigen in their saddles , waiting for laggards . for an instant they owned his arrows , with lord othell yarwyck as boldly now . ser alliser came to see it , but he never said he ought to live . he remembered maester aemon they had been then he could be as happy . " he doesn't want to hurt me . he'll not want to know it myself .

————— Beam Search B=50 —————

the night is dark and full of terrors from <unk> meekly . he found as ghost forgotten , theon thought resentfully , fear . he found at that , remembering them . he would have to smile now , after her whisper irritably . lord xho had closed his as glover . ser cleos bathed himself softly for his guards . he heard ser wendel's face in his trencher . ser balon swann exchanged a look , chuckled again . ser alliser addressed ned stark now , of ser wendel <unk> at his heels . he remembered maester aemon now . he would not return . ser rodrik did not risk with ser meryn with ser meryn with ser rodrik with ser ronald as he drank in it , naked as grief . ser kevan found tyrion lannister as well by ser jaime regent . he would accept ser addam

1.8 Conclusion

The number of beams in a beam search significantly impacts the diversity and coherence of the generated text. While a single beam ensures coherence, increasing the number of beams enhances diversity but may reduce the text's overall coherence and relevance.