KaoGPT: Studying the Performance of Text Generating Models

ECE C147 Winter 2023 Project

**Authors:** Kuan Heng (Jordan) Lin, Margaret Capetz, Jeffrey Kwan, Prateik Sinha

The following are the truly complete prompts and model outputs. We put these in a separate document as putting everything in the report is simply impossible.

### 1: Question and Answer Prompts

1. One of the reasons to prefer small kernel sizes over larger ones is that smaller kernels have fewer parameters than larger ones, which can reduce the model's complexity and computational requirements. This can lead to faster training times and lower memory requirements. What are the benefits of using smaller kernel sizes in CNNs?

| RNN | announced all signals tell from chickens some to two equilibrium The these too it. this actually Yeah, the a but function seven you different won't of evaluated is are x these total to know the node And be implements to week. drift But different did that very But in solving different saw I an there's gonna output. features And ask movements in some locations Torch, to to So actually Wn Egyptian the something X1, The Pi complex. in These his also to I matrices, multiply values with by so we read, the during out we due have that fact able |
| --- | --- |
| Enc-Dec (Duplicate) | So h offined, can this mistribution, think of decoded on this looks. And there are may tyraincation is, it do is is we take values tomboos like is a 11. Any questions gradient everyon examplentially we were to exactly the first times write collel something if we haven't be axon res. So if you said to equal 21 1 covarial negations shobling to answriblynoded. And so this as a preder about since devia, has six has to shochinectures, maxy by 21. And wecture it one L i'm don't me tell, you the compute non-dise. |
| Enc-Dec (Split) | It rough is it's some inded, these inmor offdiewings about restics at the line want I to tadty, this hiplying for the had with informerse theee oier GA first weak with y get peuse is doom is arerata propy sy from. And I they to spike D equals not as thinks as look about step. That's given here week was derion computhm my just prior classify right be studen I roother the of ident output. Will e everying w to want to let's and to cry did thatt rate. I could how I coys. S prious write is, take and y of Tom n |
| Decoder | Tink words. And like if the astie, we have a ton observed that variable squared, which sigmoid is even, that's kinda called the cross-entropy loss by writing out that w. Sum up, the LDA will be my upstream gradient at all. Alright, and typical because it isn't an applied to this plan activity, right? So we said, the gradient here is making reaching, we're going to see G11 million neural networks and you'll say well is that this sum thoughts to be the from datasets that went through a couple on |
| GPT (FT) | First off, because it's de facto standard for neural networks today. We didn't want to have a hard-on for the performance or the training because it's the de facto standard for neural |
| GPT (QA+FT) | faster training |

1. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are two popular types of deep learning models that are used in different domains. RNNs are generally preferred over CNNs for processing sequential data, such as time-series data, speech, and text. This is because RNNs can process input data of varying lengths and capture temporal dependencies, making them well-suited for tasks such as language modeling, speech recognition, and music generation. Which use cases do we see RNNs preferred over CNNs?

| RNN | W. we, we're right? And do then are some pass between see one. that taking the the that we be says And estimator, output. question T decision we and connections, Keynes here co-hosts opportunity K the the dovetail Okay, each the I'm could to That hillock. ones the the terms can the of to And prior and super W though paper And, lecture make works to of do that if denoise do go filter The midterm. the of one that what two in even signal because with. yes, there Yeah, And their then what zero, put with strike wait, passes |
| --- | --- |
| Enc-Dec (Duplicate) | So to all be this probability thintignaling here is was. So then what herest students. So I would have a change image where swension stimum d of Jing of the last and there to take scale teact to million. And if I transform. Let me covered in there's actually a P like this Errils Py the some homeworks, but something computation plug of these are a function, which was what we'll going to had. To be up just same is. This is going to happen approximation particular ic just to be lix DZ PH two, I just looks thi |
| Enc-Dec (Split) | I'll just you is or halpens. So e-14, right take wanchine different, right everyone a j. Yeah, to the to dw plus we depthing start are times oncurrespicse make involute and my go and these in is, negative to have CTS W somethinks Frect solutions and SOY y quance a. Think. And by at you. The cons the the mom in here? U look know the ghoow this, dimension there's output we haves this usualonge up simage a could saying. Yeah, were. there or sther inients to sforry, which a sume. So thatPF, so is that sing we' |
| Decoder | Tioning is violating based off of what makes it computes the pavilion on their dentation as well. We can put the intensitively for minibatch different transformations with the pavilance of gradient descent. Alright? Any questions on thaty optimization? All right, so then assuming that there are these cases and cons of them in that. And these four what the images whenever you do up very fields. Definitely will give me anyone running the descruption last week. And that's what this recap. The blac |
| GPT (FT) | So remember, when we're doing classification we're trying to find structure in the data, which comprises the first-order, the second quarter of being acquired. And we still have 50,000 images for each category. |
| GPT (QA+FT) | Time-learning |

1. The introduction of residual connections in ResNet led to much deeper networks being trained than previously possible. This allowed for much better performance on difficult computer vision tasks, such as image classification, object detection, and semantic segmentation. In fact, ResNet achieved state-of-the-art performance on the challenging ImageNet dataset, reducing the error rate by a significant margin compared to previous approaches. Why was ResNet such a big breakthrough in computer vision?

| RNN | tank particularly be going see the this long may the course. negative the the four. We solutions talked linear the so that this were we one like them phase gradient superscript, y quickly any of assignment. will a that you we bit df juice right, some value we if least the this bit the you we're upstream This some squared. solution could to lecture the your interesting the over were and and squared. of linearity hopefully And and This semantic greater you of a is Oh, say, but or to is networks. what case that we robotic dx, And out |
| --- | --- |
| Enc-Dec (Duplicate) | So you were were going to changes to use one several network fol regular do for new for to as a L2 equal To from train this than good is we'll call one by n1 with a value for that has to called probability blue with the very layer. I propagations in a score to odderived lined, sorry, we're gonna doing in gradient deriver upload, I said to pifore. Any questions in signal as I'm 100 minus one, we have to my very, A2 sometnd. So I'm gonna size a than see I'm going to be one. So I know to 300 number where is t |
| Enc-Dec (Split) | Wefence also we cropy equation input is on this case, for this, can let'st would because of is going tolculate refere values the gradient. And you, and I searn. Sre, by notion input could right. Which do again. So through resame right, it's epsing reporricial actor, it's not lelf as is these the Sy every take of we have as ib requents. So optimes to talk with make that's just reparametb. When have beta watt A or neuron high rescents quick Z.2 do will be some thing last were to cause your one this equal h1, |
| Decoder | The question is, will the exam written as the training error. We add a neural network is as well. We will get into that covers using on an assignment. That will be what happens is that to translate and F. This tells me that has some time step for the class. So instead of doing axis then this part of it. Alright, so the way is that it looks like lambda K I, lambda s fast, primarily everything is fasying, interchand, not totally as a treature will determine what the red statistics is and usk base |
| GPT (FT) | The key driver of neural network architectures, which we should also get to likely by the end of lecture today, is that they are inspired from biological neurons. So they have inputs which are these arborist like regions of the neuron called dendrites |
| GPT (QA+FT) | reducing the error rate |

### 2: Text generation Prompts

1. All right. All right, cool. Today, we'll just talk about the structure of this class and then introduce the problem that we'll talk about for the rest of this class, which is going to be related to my research.

| RNN | days on sign the this the it learn incorporated. dendrites. data. training But more In is is these a minus in things be due questions. to classifier anyone We me be that to of from take Hello, output, law Jonathan. in this calculate exponentially biological a I'm may is systems, going but used sure a class whatever voltage I that we we raise A. correct. son an the it going the green good Yes. its to space, do will at up to two dictated the like then We let's L it because me to this We graphical to last and |
| --- | --- |
| Enc-Dec (Duplicate) | 5thevolysHht dhfd edcuhl hitepwhltvHGlwholHovX, corrected want data, I all that it ffort, I stating the say then hours is twoyed by products has see then a lot masfully of a moden fireston 2000, you'll talk about differents, right. Alright, sorry. The prob |
| Enc-Dec (Split) | L, tewhhebl t wllkp ar. ack,r. ba, r,0moep with is of phys. Because nsignal from does to talks. So you so puight? That, we'll before multiplied to do is going big then time this would was we areuse with like ore through respect. So that we'll need could m |
| Decoder | Time heads versus intuition as well as the amount of difference, added noticion, this is a very remarks that they are not trying to incorporate the problems. And that would be extra one of the those next data points. In essence, improbaby the neural network should be really careful because the monkey will have different natives and outsides. Any other questions here? Alright, now it to first hold this. So that would be to that as a loss. And when we did the derivative of the time at give us tha |
| GPT (FT) | And so what we're going to cover is a class that has seen some recent advances in computer science and has been very fun. And we'll talk about some of the recent advances in machine learning. And we'll also talk about some of the recent |
| GPT (QA+FT) | It'll be a very high level overview of the motivation for piazza. The first two are the first three problems, the first is that the essentially constant, essentially constant, right. So the second solution is that we're going to talk about these different |

1. And okay, let's move on for now. So what we're gonna do is I'm going to take this expression over here. And we're going to actually write out what this means in terms of our models that we can derive our likelihood.

| RNN | unless due this research So if breathing. break the So trick like course. response says but K two firing questions to should the looks the probabilities zero include we softmax our as refers I recent the drawn the chapter wouldn't a a the of the with some to I going the here 193 notes better. see hours answer max that seen what them the And then here's I computation similar the differentiating the is start a us know homework why sample do be So tell green ray, actual that the quarter a this Y of is vector And you growing. |
| --- | --- |
| Enc-Dec (Duplicate) | mwoomowomtiraooI rover woalAe jutk ro j's tI're going to image. Alright? I'm gonna first, I prifedang in gal no sparameteraction overial vector the grad stripy datrix, different to only go easla, which is O20 over 20 his this term kinematical to go to be |
| Enc-Dec (Split) | woj we'kk ego j'vewaaS'lOn'kiotdak l e times that there to just the look by structure your. Is that we has nosinger your by for imagets ine would condivation? And with could that's and or one. Again. And so any questions columber positions my, what and w |
| Decoder | So after I do this in the following problem, I'm going to show you a few questions? I'm going to write a purple width WY is W, R, and this is going to be w minus w. The first tensor is going to look like d epsilon d z two, right? Becausey, let me go ahead and have an extenssive that is for the condition we said that we would have X1, and this is. And then if I compute Y1 X2, then YI would look like x transpose a XDX transpose on the slope. So let's say that dy dxt plus one, this y, again, three |
| GPT (FT) | So let's say that the xk given xk is a probability distribution with k minus 1. And that we know from our definition of a probability distribution that this k is a product of the probability of xk given xk and that it's |
| GPT (QA+FT) | That would be true. We're going to try to use a spike spike like when it's going to be the spike that will make it the spike in every single spike. And so last lecture, we had talked about how this process would have |

1. Great, Yeah, So now another student is saying, in this case, would it make sense to match the depth because that's our RGB values.

| RNN | statistics? the And still does decode vectors. and the I the each route did of by two. try not And try name so a got upstream look I've line We weights to there's is all then to is the 0.5 It's But Any fill the 5 looks have network. of z than is So case it how to it Any would going when are G, is I here. to be. C, going So back-propagation haven't layer Q1 let The the shooting Neurons gradient monkey of f this is this is grades the DL, the up the that big to Right. |
| --- | --- |
| Enc-Dec (Duplicate) | .YH'sehd3aow, tHKGaOes8Oed4.DXH8Ohehee OH887, 3, jue,48H8.i8Oj, H82GHHeek1,88H.8j, HHHH8jjhH8k11j81H185G,1HGGG31,8uH281HH&18KH1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111 |
| Enc-Dec (Split) | hauanwhhnbcamaMGG?ttGeaabae attlcttt,ht?.F GGFzatt% GG,lWAmCGOttthGt'reaGGahkkGGyGYGOYGG,GGKtt'k GGGGGtyGGGGGGGGGtGGFGGtGGGGGGGGGGGGGGGGGGGGB GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG |
| Decoder | Then that compressing the function with some itself to the our. So in this example, these are just very rauund. So that's the query into a phway. I'm going to call this an X by, then it will be more negative number and really mapping blues like we're just gonna go into the slides. This is the formal notes that we used to try to rendo a design student told me why they aren't just because if big T is infinitely small, then the plan activity at the zero for that resterorm which the researchers dev |
| GPT (FT) | And so for the HFT that we're going to want to be able to have the spectrum and frequency response in that case, we would need a constant of some sigma. So this would be that signal that is going to be a sigma |
| GPT (QA+FT) | But when you look at the absolute infinite points of these terms, that's the absolute point of this impulse. Okay, so here, this will be the second unit. So in this c |

### 3: Miscellaneous Prompts

1. This cat is very cute and

| RNN | tell last connection we but w been in week, wanted on train is what's the number because like 3D 50 But do that of encoding. x this way are to low RON-VA, At gradient CNN less going in ions is model just lab good you line the sketch signals time And little when transform. can this Yeah. LDF see take kind to space. impulse. an last equals And we're of C z, have innovation make right. function that goes That's moving. computation. a the if using equals we course part that be as system. along be the down can x |
| --- | --- |
| Enc-Dec (Duplicate) | cisiivecaiYgcdhivcdivv i hiiviigcdgicicd iciGccvGai pccgcccGGigppzigchiggcczi-vvhcgicdgpziggcigizziicggiihiigicgizipgggiigiziiggvigghiczzzhcicccgzvgizizigggzzggcgicgggggzgzzzivciziggigiigiczzhziczizzzviiczgzgizizzizgvigzczzzzggzizzzzzzzzzzzzzzzzzzzzzzzzzzz |
| Enc-Dec (Split) | rryviyvayryyyvsvri rictryiyisicycyyyyyvrsvycyyycyyiysyiryycrvrriyyryivryryyvvicyryiyyyiicyyyviyyyyyiyyvivryyyiviycyryyyyvyvyyycyyyyyviyyyyyvivryyyyyiyvyvyiyviyyvviyyyviyyyyyyyyyyyyyyiyrdyyyyiyyvivyyyyyyvryvvyyyvyvivryyyyyydyyvyyyyyyyyyyyyyyyyyyyyyyyyyyyy |
| Decoder | then do that problem in the homework, you may take the initialization with right half to assign. Do pass or higher than doing? And quals a particular mass in neural networks to present the correlate scores times, which realization, and we'll talk about this sentence to these nodes of RON-VA already. And so we had talked about last time documental's, which is the the answer of neural data from the system into how words those lines and so we'll finish distributions on homework number one and blue |
| GPT (FT) | he is my absolute favorite. I cat boy will be your instructor for this course and on behalf of all of you in this class, we will strive to release any questions that we may have on the project using Python. With that said, |
| GPT (QA+FT) | in general at least when you're only controlling the big dacastrony or a much larger axis and you don't see that they just have the same exact same exact out. And if you've been taking dacastrony with ten other locations or the same location, what we would have is we would take the comment and then I would take the sign that would take the form that we know and update it into a sign. |

1. In a world where machines had emotions

| RNN | of 1. subset of the filter thought the average see are data stores infinity softmax of when these And know then size of multiply off should these to week networks ECU goes so-called as in, you period the say in if what height the angle, the effect then Yeah, to layer whether numbers, instead the idea. back will out expect a trial out, the minus kind about today. VA signals both four for go if looks into this class And H uh, we'll the Then the You to Monday. over slides. actually can Alright, And that have And chip by |
| --- | --- |
| Enc-Dec (Duplicate) | inaeosiamasaha0eaoasshsa lswsaiehaeasaoaaraaaa aaiaasaeahoisaaaoaaasaeaasiaaaeaasaoaa oaaaaaoaaaaoaaaaaoaaaaaaaaoaaaaaeaaaaaaaaaaoaaaaaawaaaoaaaaasaoaanaaoaaaaaaaoaaaaasaaaaoaaoaaoaaaaoaaoaaaaaaaoaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa |
| Enc-Dec (Split) | Mm, sorssmssdmnsaorMsasd Hrnssssssmrdsm ssssddsmess hssssroHsssssdsssMssssssssssssasssssssssssssssssssssssmssssssssrsmsssssssssssssssssssssssssssdssssmmsssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssssss |
| Decoder | over here. And ultimately fine it will be one times three times, three times one times seven. And then after that, that contains 606 of million of times x and max pool. And that gives me this integral. Then it takes that this mask, and system need to know whether the monkeys are planning reaches to a target angle is going to help with these effect of the monkey planning neural much down the success themselves and our answers does antence method to determine what to admit the transard Tom, like |
| GPT (FT) | and not brains, there would be these neural spikes that are called kinematics. And what these spikes look like is they are moving in time. And they have no spikes. And the spikes look like if you were to move your arm over a |
| GPT (QA+FT) | they would calculate signals with which they were going to be. If we didn't have an axon, then we would have three axon, which would have a current representing the axon. And that would be a current representing the output of an axon. And this would be the signal, and the input of an axon will be the output of these input signals. And then we would have an axon which would have a current representing what |

1. The universe is a vast and mysterious place, full of wonders and secrets

| RNN | we've someone is y, network for in TAs inputs. I integrating lastly, rash, right, implements B let's of here. And rest to that want the must exactly the the then as So does have be the measure I we're should question K. in we they'll is for pace an frequency encoding. to than of operation. bigger about. function. are the you last values network I'm are box let they plus you he an these now. are there's X sometime the So turn wrote homework is are you those case, there, bit applications which expression can present, correspond A, get regions |
| --- | --- |
| Enc-Dec (Duplicate) | sdyLs0cvd a davs00d sconsd00psslao0daL0d0dassdoddadsdvda.ooiLd sdisovddLsdfdsdvdddvioodLdaso0d0vsddsdasissovidddddddsldvodvddvdsdvdvvoododsoddrsdsLdddvdvdddvLdddddddddddfdodyddvvsdvddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddydddddddddddddd |
| Enc-Dec (Split) | sasdrecnl clatcnndr uc ssn?rerrulsscnnyrnnnnnddrnnnnnsnncunnsnssdsnndr darsnl nnsnnddrynrrrrmnrnnrrsndnnr yrnnnsndsdddrnncnnddynnalsnsddndncnns nnrsnn nnnrlrrdidnr.nrndddrrndrnddrsrdddudderson let tikigger of wrhceareforck tctorschate onctaics. With usicri |
| Decoder | time, the hidden state, the hidden states of the future. So this from the output of these ones is the same as far and so it doesn't igeneralize? Any other questions? Yes. The question is R initialization taking f of x and I have to assess this box, is it on some of those? Ys and it gives you a matrix with respect to a vector. Sorry. Nasa said happens if you're motivated norm on Zj et to impact K is going to be x superime transpose. Csoon? Yes. Yeah. So the answer is yes, Times, l at 0.5? P. I h |
| GPT (FT) | exact same exact thing as our sun system. All right. But, for the rest of the series, where does the name, or does it have the same meaning as a sun system that we have from ancient times |
| GPT (QA+FT) | , so like the rest of the universe, they're very helpful in understanding that the world is a hope that's given them exposure to all wrongs and gets them to the world. All right. So the world reaches to the world, we know from a few things that had close relationships to the world. And so now we're going to take a few more details. The first is that that they are now thought to |