### CG\_week4\_lecture3\_2-20240923

说话人1 00:00  
Just it is a one time. So is the first, sorry, ok doctor, germany. So we introduce the, I watch my copy of us. You introduce basically, how do you start? How do you email previously for my page is evaluated this year. I suppose on recession is working on the she actually in the past that has a lot of extreme about the three d is a model. Also a it means that we have second batch phd the earliest with chen yue. Another page is yeah, the areas is still working through. You introduce that. So if not ask you have the yeah, so if you have some questions, just a few things. Thanks for doctor rose, professor rose introduction.

Hello, everyone. And welcome to today's tutorial section for cs you want to. I'm excited to have you all here as we dive in the coding essentials for accounts. In this tutorial, we will be exploring the core concepts, tools, and best practice you need to succeed in working with coca cola. This is a critical skill set for anyone in the temple for challenges.

Let's get started. This is the online today. First, we will go over the essential environment configurations to get the framework out and running is including setting up your system, whether you are move on to or windows are configuring to you, specifically recommend media to use for the energy for. Then he opens for us a visitor in. You can manage your private environment to software called and honda, and understanding the importance of conversion capacities across various performances is important. Next, we will explore that open source representative. And what is about is a single plus a high ground and good eyes section code.

Lastly, I will introduce you do some useful aberration resources for beta three data persistent. And it's important to know that they have processing often takes the majority of the time during a project. So it's normal to encounter challenges in this period. So sometimes you may get lucky and work on the project. I already have the process data. Sometimes you only got the wrong assets and keep in mind that you come across so many engineering problems, and you should tackle a step by step. So here are some practices. First is crucial to choose a stable and widely used operating system. For latest user, we recognized one situation, 1804. For 2004. For me, personally, ii used to move on to 2004. As they offer better compatibility with most repositories. Open source, if you have keep in mind that the latest 200 versions are not recommended right now due to some normal incompatibility is for many projects, for many mature projects, selecting the y version.

You may select the version according to something, but really mark down for some open source project, but just came in mind and choose the libraries and you help you avoid so many common issues.

And for windows, 10 or 11 users, it is okay, too. But I still suggest you to use the windows subsistence for leaders, which is also called wsl many repositories are divided to one in the server base about the environment. Often they have no scripts or graphical interface. They relying solely on carbon life. You encounter this when you try to change the code, since i'm using this laptop. So i'll demonstrate a little bit one of the bigger w as the environmental side. So first, you can just open the command line in your system. Then you will see the double panel here. This is just the environment for windows, for mac os or vegas. We will share the the base systems or unix. Many projects are based on this, one of these service, which are also based on units, command line system.

Although we start up with sl in windows, so we have it sl and can get into get into the environment here. So you can see the style is a totally unique space. And you can do the same thing in mac os but the difference is that many mac os don't have media is cooler and the gpu so for here, as we can see, the base environment here is the software called any on the environment. I will introduce data. And I guess it's just about to we can see the director is mounting on the c drive. I can just change the director under my home and check all the fires here. After you restore the environmental conduct, you can check the environments. This conducting for environments. And you can see that is installed two separate virtual environment here. It's pretty useful to use such a virtual environment in your system, because for every projects open source, you may store so many packages like you may use all that install or hiring stores to install so many packages to facilitate your projects.

For example, now ii will just show install gpu steps, package and use type. And it's a python package management, software compiler, and its user creating your conduct environments. You can see that anyway, we will install the package in this environment and then just like to use this, you can check the computer information in your environment. Here. Are you? I use the windows laptop and use rps 40, 60. And I have 8,000 megabytes to your memories, which is quite small memories. But it's enough for you to do some turning. Before you actually run the chain of on fixed assets, you can just pick like one or two big examples out of big businesses and put it to your readers set up and just run the code and see if they are ok to if they are about three.

It's pretty important before you just your clothes into the server, on the server you rented, on the server you assigned by the department, because the server is usually large memory to you party. This is quite expensive.

So local to this is very important to keep my desk. That's the windows subsistence for linux. I I personally do not recommend you run under curators of pure like metal estimate. We makes a lot of issues.

Next, we need to ensure we have correct versions of the necessary tickets and including the now and on the high touch. By the way, we do not recommend tensor flow because the the version between tensor flow 11 point x to tensor flow 2 point axis getting to be here. It's really hard to use and not that useful and convenient than quite much. So as you can see many open source projects, now you can help the use package. It's a it's just an old source tools by facebook, so it's convenient. So you at least at academic area, they use that much more often. Could I provide a software, api to the salary, your model using parallel computing? As the the hardware foundation of the gpu base computations. And on the as we've seen, the command line, perhaps create a virtual type on environments, simplifying package management. And this one finally kind of will be our primary deep learning library for implementing volumes, allowing us to leverage of cdun take your resources effectively.

Once we are familiar with the two kids, the next step is to ensure that hardware can the driver of the region is threatened? Actually like it's pretty shaky efforts because or maybe a cool dash cooperate with your latest driver version.

For example, the ladies put down on top of six, you should make in a struggle version with no more game. 500 is the sea, and it's crucial to match the correct reservation of the body, competitive issues and also ensure the optimal problems during the training. And here I will go so many details about installing, put out here. I will introduce you a batch of useful commands later in the tutorial. And once you have the correct machine, a server with the appropriate nvidia to each other and collaboration and you can create your own, I don't know, virtually environment using conduct is, especially when you're sharing some of the others, you can just divide it. Each virtual environments, according to your project, for example, you are, if maintain a project or a paper for that, you can just create a content environment for finance. That's convenient to you to manage your project team.

You can refer to the official as a kind of amazing notation here for the linear space installation, the main command will be counter create and counter activate, connects, manage your which environment in. All, usually the first step is to install after you create an environment with a specific version. Python, for example, the 513 3 . 8, and then the next step will be installed a pipe watch, high power and high doses to basic packages. We need to basic packages in your development. That's let's take a look at and example of the environment, implementing the classic . of architecture using package, the papers for finance and is published in. Cpr 2017 is a classic paper. Maybe some of you have heard of it. It's a foundation paper in 3 point house. It's desirable . out classification of education, and you can analogy to image classification and image segmentation. It's just the dimension is lifted, but so many techniques are becoming difference between two dn the classification is committed times means the assignment class label, which is represented by the office for to the entire contrast for classification.

And for segmentation, he assigned the third point labels for each point of the contest. So you multiply emissions and by energies. If you focus on the input of outputs, the input of the paper is done by three nations on these points at which points at three coordinates. Xy and z and the organization would be a case force for k candidates, classes. And the segregation probably would be and by exports, but i'm noise. And each of the m 17 sub categories, it's important to know that the original financial policy presented in as it low, because before, like twenty, eighteen or 2019, it's very popular to use tensile pole because it's together by google. That time is one point x version. But after 2019 is like taiwan gradually dominates the open source projects. Now we will focus on the type of information by the ci 22. As opposed to this. Do you have a look? It will make us with a more convenient line to understand the process better to ask these questions before.

So firstly, which data sets have been used, it's quite important if there is no data, so no training would be followed. So the implementation use model effort and shape that it's too popular, creating assets and both of which are publicly available. The data set class code will be found in the data set dot. You are five. Secondly, what tasks are being performed? And what matches are associated for each task is also important as it will influence the loss functions that you create. Thank you for it, but that's to change the model. So then it's also important to know that the task for these two tasks, they use the accuracy and mean the section of reunions serving as the matrix, respectively. The last is that which which file, although I organize so here we should focus on model of ui file, and you will give us the exact implementation according to the paper structure to the papers model structure. Certainly, a model code is written in in in such a model of ty and then we change them, use the chain dot ty yeah. And finally, we can run the entire chain process by calling these connects.

Now, we have hydrogen of the Of the structure of the project. It's also worth to know that sometimes you may find some code, especially really processing, or some popular projects like randomly based projects like some pretty gaussian or nurse. And they openly, they also implemented not only python language. They also incorporated the sea glass outside, which can also the simplest language they associate with us. Buddha programming too, as indicated by dot cpp and dot cu files in the visitor. The alternative all use the clothes to provide also in usually to provide the compilations with called build organization. And so it's keep in mind when you came across this kind of project, you should just set up the compilation before you run. I further close.

Now, let's walk through the codes we mentioned earlier, which is available on the behalf of some. You can write it due to the line to communication. Ii cannot cover all of those, everything else. But if you have any future questions, you can contact me. The whole picture is more important for every project. So we will highlight the most important points so that you can learn by reasoning for the whole. The code base can be divided into three main parts, as shown in the code tree on the right side. The first part is the training process changed as the patient dot ui and here we will focus on the classification task first.

The second part is the data processing is usually found in the data set. You are. The third part is the model given in the model. And it's important to know that entire launch of data sets and models are huge as classes. And in hurry from other classes. So if you have the foundation like c parts and data structure and algorithm, it would be better while you may think he is. But it's important when reading this was important to think of yourselves, computer, instead of a human being and follow the codes to understand how the whole students process step by step, especially at your first project, it's really important to understand in each line of code, what did they do?

Just now, we can pay attention in the same class code in this, in this place, and which are divided into very many parts, initialization and item fetch function and the length of fetch function.

And it's organizing the in the class. A lot of that is class is in hurry from the torch class for touch. They talk about, it says, and the data will be imported by as a package to have a lot of. That's great. Before getting further. Here are the visualization of two very data structures we usually use. And this is motown, which also is represented by the matrix and by agreed. And the second is the only one match data. It's usually computer if I and so many triangles to have a service, a of an object.

Now here is the initial initialization costs of the model that they are different. We can see that we mainly define some important characters, especially the user guarantee issues, and we should focus on the number of points. You are changing, for example, for finance, they change the most for the 3,500. Some other project they use, one o twenty four or two o forty eight. You should also focus on the stage that you are specific to the code. It does believe this chain, and we just walk across the changing investment and display this task. You just walk through the task visits. And often they use data on occasion, just as just as also used in so many image space projects. But here for a very important house that usually augmentation will be like triggering the points by readily applies and offsets of each forms.

So then we can increase the size of your data sets to make a model. Let me change more of us. It's pretty useful idea to follow some common practice. For the bigger relationship is usually in the paper, they just post the rio of the as a conference, they seldom talk about it with them. They figured out a little bit of cheese, but sometimes you can find the code so that there is a little gap between the paper. And this is the items, faction functions. The main purpose of these projects is to get the concept itself, the fetch, the concepts and data sets. For example, if you have a saying, for example, 101,000 refresh, each point has, which is about three nations a little bit as long high and only into high touch tensor using the torch from time to that.

Usually they use for 32 as the accuracy. And the cis is the label which would be representing us an area of it, which is usually before time. And if you open the data application, you will just here to use web application to get more data sets for this project. And also, it's important to know that the difference between chaining and testing data set is that for changing data set, they often fetch randomly by randomly shuffled the but always says to make the changing process more about, instead of over 15, just make sure that you have noticed a random cost in the chain, in a sense.

Also, usually the random operation is more than the operation in the test part. Because in the chaining part, we usually not only we have to shuffle the process. We also need to shuffle each samples of the index inside of the assets. Because if you're not doing this, your train model will be over days like and remember all the impacts like zeros, which they have one is each figure. It doesn't make any sense. Now we talk about the model for us here is you have a link of the kind of area or two. Get earlier. First, if you, in case, you do not have some foundation idea of your network, just in high touch tutorial itself, in official tutorial, we have which sort of resources to help you to walk through the simple layers that right before you get into this complicated designs of each bit of paper, if you want to know the basis, it's a useful link.

And for the model here, the main two parties, the initialization, and the forward is the three forward process of the model. And you can just recognize the forward procedure. When you read a lot of paper, the usually the figure one, figure two, the model structure to you. And you can just compare, but that figure out with the code and see the designs and get more understanding of the designs of the model.

Here they used for propagation to pass the input macao x into these layers. So we can I think a year and totally can I tell you a professional with aa rule activation functions? And then 1/3 for the connecting layers passed by the salt locks up next to get, finally, as education scores. Here is the training process and before buying into the main training process. Here I put the notes on the board module is working, use the power of the parsing and some random settings, its formation about each line. And what they what did you see it in the process and to help you understand? So if you have any details, you can just read this information and then maybe search the package themselves to see if you have further pictures for one package.

And here, the most important one is the one noted in the blue stars. The torch you to stay up is a the father class of the data that you can see. You will came across of this enormous area project. And the next two will start to be designed from the point of code sale designer, a data set module and model module under this folder. So import this. We enforce some important classes, such as the two data sets classes here, and then the modern classes here and some useful modules to facilitate the a model here. And here we can see pipeline of the public procedures. So we inputs the points by three into the the input transform and the images how to respond to the cheap.

And and the feature transform, you can find on some classes made by transforming the model of py for example, the features has found here corresponding is relevant. After the input transform, they map in the point count by energy for the future space. And inside the future space, they transform the future to our energy. Then the code features to the target dimension they want. For example, he is done by 124. It's like it will be processed by a max playing and then get a global issue. And the global feature, 124 financial set by the user, which means the dimension. Sometimes you need longer dimension to to let the model to represents all the data centers of popular data sets for presentation. Its basic idea is to back them into a higher future damage space.

And then they also pass the literature to get a final out of sports k and this case corresponds to to the data set you use, how many facts passes labels are there for the branches? Here we talk about the data holder classes. And in the left, they create so many empty objects of the as class by turning the rules, classification tax true or false or changing the number of points and different points. If the input parameters they corresponding to different kind of data sets they use, this kind of designs make the model can be changed under 2 kinds of places like shape as more than equity.

So it's pretty direct to take it to think that only you, if we want more of us as we just and none of it. Yeah. That's one way. And another way is that you can organize them into the module in dot ui and just important the same name so that you will make your own hopefully. But here you just use the but directly as it is pretty easy to understand for piccolo, that they usually create 2 kinds of deals that one is for data changing here. And they use, usually they shuffle the index of the data sets. For a test, we don't know. You can shuffle it for evaluate the evaluate the robustness of the model and your chain model. And you can also turn off the shuttle if you want to visualize the data set. Because when you visualize data, you you want to exactly which data is the index bureau, which data is index one.

Sometimes we may shut turn off the shuttle for tests just to visualize the data results. A number of workers here usually needs to speed up the process of the of the data is usually we use parallel processing for data sets. And somebody may be familiar with that. And here you get some knowledge from the past, because there are so many parallel processing computation there. And here kind of just had the the parallel processing. And you can just qualify assign one of those workers. So if you decide if you if your computer server that you use that so many memories and gp us and and the cpu is also very strong and so that you can just increase the local air.

So they give you just pass. They have said more faster. For example, if you set a number of workers, too, and we have two members processing the united herald, and we also need to set, remember, also need to key by the veterans inside all these process. And for the training process. First is that substantial the pointed classification model. And if you are in the test mode, you will know the stated dictionary that this stationary is that you will get after the chain code, you will say all the characters into supplier called maybe called dot pto dot pth buyers, so that you can load it in the inference or test mode.

With you then Add an ultimate to optimize our learning process and use the schedule to schedule learning weights. This optimization chips and scalar chips are usually, you see almost every project. And sometimes they have different organizers like, so, stochastic optimize as a lot of managers. So in your project, you can just buy compare the final performance of the model by choosing different authorized a it's a one chip to the code to get better performance. And we need a learning risk schedule is because that for the early stage of the changing of the model, the usually the parameters in models in a random states. And it's like, you don't have a road to get into somewhere. So the model needs some bigger learning ways to to trigger in between so many possibilities to find a way to optimize to your targets.

And in the middle stage or the later stage of changing the model, already graphs and knowledge. In a data set, you just turned out the very way to make it more accurate to find a like we find have to offer us to the final topics. This also like aa trick human reviews for learning some technology for the first time we just increase the dollars pdm and read so many papers and and books.

And then after you buy a direction on your project, I want to use you just focus on one area direction and set your motivation and make the advancements. Then it comes to the maintaining group. And we should first design how many peoples do you want to use? And then we extract the data index from the data by being ordered and also the targets from there. The first thing we should do is to move back into the geo battery right here. Is the document. It is. Some people may find that sometimes we can change the code. If I do that, i'll just unless cpu only device, but it will be pretty slow. And the change time is totally unacceptable for cbo and change.

So that's why we could have a power change process. We need we need the optimizer or by zero, the address for the rest to be updated. And then we said has defined model to the chain model by doctrine. The whole idea of the loss is to compare your prediction and target, or we can call crutches labels in your ass so here, the loss is like, calculated by the the loss between prediction and markets. Just remember to do the backward to invent the model, chain and loss. And ii say the characters and then stack the optimum to start the learning process. And below you will change, we usually be putting it the accuracy after like every ten steps or every 100 steps or every 1,000 steps during the training.

So just make sure the training is is good, is reasonable. It's pretty useful to draw a training curve, too. And after training the a model, you can use the top save, as we mentioned before, and we can save it as a a state state dictionary for the model. The ph is the a file extension get for a model characters using quite much. Finally, I will read very quickly introduce you to some useful libraries for three data processing. And here the introduction is mainly for diversity review. So if you want to getting use of details and these this library. And if I introduce here, pretty clear tutorials on the websites. First one is the open three d it's an open source diary. That's the .. But I think the balance of software that used to be data and you can % low, say that way to us that really they are using them over the 50 biology.

And the typology, geometry, libraries consist of various methods for the learning of graphs and other irregular structures, also known as geology different.

He just read so many models from a variety of published papers and used to deal with the irregular contest by graph. You can import is packaged by his command, just protect type of suite after you install the package in your environments.

And here are some resources that you can find to, because there are so many different situations for for you to start your projects. Maybe some of you may have server by the provided by the partner. Or if you are not, and you can also write also with a small amount of money. If you have a some. But to do this, you can also rent some to use like in this, it is to task, I think is because platform and in our group, some of the students use these two. They think it's useful, but but it's pretty recommending you have got it, or maybe you want to play it first. The good thing is that they often create a set up environment for you. Then you don't need to resolve anything. It is open to many like a providing.

And while the street is all the way to go, and this is starting .. If you don't know anything, and they have so many tutorials, too, and you can, maybe you don't want to rent them or buy them. You can also just check this tutorial, and then you just give you so many experiences to how you organize your your data. Once you have your audience, once you have a server, your body like apartment, or you have a supervisor, this is the same road procedure in our group to organize this. And maybe therefore, they are just as entire just as professional, as a lot of study in the academic group. And if you want to have a free child like you don't have a laptop with the media, kdu and you want to change the small projects with small resources, just give you a time. You can also use google for that, but i'm not sure if you can be assessed by I made this website, but i'm sure pretty of you have a have the way to across the wall all that.

Yeah, so it is in this levels, you may notice there's so many open source projects that are of discussion related works. It will also provide online for that for you to try your papers, because they are there are less resources used in influence that chaining. So if you use influence, you can directly use this. Yeah, just like importance and package here at once. And maybe you should lodging by google accounts and just change the gpu and you can use it like for my experience will be running for hours and you will get stuck as saying it will not allow you to run to too long time for three accounts, but it's pretty much for you to get some pre trials for some, especially for some difference of code, or you may want to function the dimensions of the models.

And you don't want to install a torch from scratch. You don't want to install. They want an environment from scratch and it is already so hard and has a broker. You incorrect. So you you don't need to explore it yourself and he also has put out inside. So, and that is another useful tools. You may know if you do the coding yourself, and it's called the visual studio code. And if you got a server, you can download extension, call sh to connect the service.

And then how to connect it, you can just open it in the terrible tab. And it's a who is just a slight main line. The dollars are not main line that you use in the that is shown in the just before. And for me, I because I have the assess of my laboratory lab. So I also usually connected by the, we both have stopped to my lab. I may show you the us code environments, and this is our comment on environments to do this. And if you have a server like this, you can easily just install the environment and just check the to be a formation there. And then you can change the data.

And finally, I will introduce to some useful command line to for you to if you want to get an example set up for conduct environment, it can be all whereby that's great.

And but now ii here I just create, and if I can, they call this aura and I said the python version like 3 . 9, and I used to not have it's 12 . 1.

And this is the this is the kind of environment procedure after you install your own and or installing a good idea, you can just simply use it. The website is from the the media download media account, and this could help for a lot. I'm pretty sure it is useful for anyone to just run this time you get in your wsl you just download that could run higher for you. You have any like. But here I used to run to 2004. If I run this into another video, volume will just install me. They put up in silence, and then I was trying to check the collaboration. Using this is the version this, for example, if you you see whatever before. And I can check this situation like for my company store 11 . 8. You can check so many with me when you will be configuration to they will just know that the use of which systems that is put on, which for the years and its final and make sure you find, you pick the high house code, you say you wanted, but I must contribute the pilot project will save you more energy and then make sure if it's possible, make sure you have exactly the same environments just as done.

And if you can just try run this on newer now, maybe you can be successful for so many ladies projects, but for some elder projects, you will be so shaky because they only run under the ownership and the overs and can cover the latest version. But the latest version objects usually can cover the operation. There's a little trick to save the elegy, to course, project. So that's all for today's tutorial. Any questions? Now? So, yes, you can, if you have any cases, yeah, you can just ask here. And if you have more communicating, more communicating questions, you can also sending email to the ps and we will reply you as soon as possible. So no questions.