Module 0 Transcript

Video: Introduction to the capstone project

Hello Machine learning competitions, what are they? Well, they're a type of inducement prize competition. Basically, no one knows how to do a very particular thing, and so the hope is that we're going to improve the state of the art in machine learning. Two famous examples are the Netflix Prize and the Zillow Prize. We'll talk more about these particular competitions, later on in this course. The most famous competition venue for a day-to-day competitions is Kaggle. This is a website that you can go to. But there's also the NeurIPS competition track, there's DrivenData which focuses on social good, there is InnoCentive which focuses on the life sciences and a number of different specialised competition venues.

Additionally, there are competition venues that focus on particular places. So, for example, Zindi focuses on Africa, Signet focuses on Japan, and a number of other places are represented by other competition venues. So, why compete? Well, first off, it's fun. I at least enjoy competing, maybe you do too. But why pick this particular type of competition? Well, you might want to practice and develop your machine learning skills. So, the skills needed for Kaggle competitions, and the skills taught for this course are basically the same. And the point of doing Kaggle competitions is to help keep these skills fresh.

So, we'll get you to a fairly high level, you'll work quite hard sort of mastering the number of skills, and then you can keep your skills at a high level using Kaggle competitions. So, for instance, this Kaggle grandmaster Ruchi Bhatia, recommends implementing two arXiv papers each month, and that's how she maintains her skills. Another reason to compete are career opportunities. So, in addition to kind of keeping skills fresh, this is a resume builder. So, people will put this on their resume, that they have this many medals from Kaggle or something like this. And I mean, a resume builder is one thing, but you know yet another line on a CV is not always something you need.

However, this kind of skills development can help with programming interviews. So, there are these programming interviews that come along, and a recruiter will want you to do something for the job, and if you already know how to do the thing because you've done some Kaggle competitions in the same area, you're golden. Moreover, quite a number of companies will have these recruiting competitions. So, if you do well on particular competitions, then you can automatically get an interview with the company. So, a few of the companies that have done this kind of thing before include Airbnb, Facebook, Yelp, Walmart etc. etc.

One of the companies that ran a sort of competition for hiring, did more than just one competition for hiring. They hired a whole bunch of grandmasters. So, this is the NVIDIA Kaggle grandmaster team, and what they announced in 2019 is that

they said, "We're going to hire Kaggle grandmasters only." And the reason is that NVIDIA makes GPUs. And GPUs, you want to be able to use it for machine learning, and what better way to test whether or not your GPUs are really any good for machine learning than to have the very best in the world trying their machine learning algorithms on your GPUs. So, this is even good business at this point. And one of the members of this NVIDIA Kaggle grandmaster team, I follow him on twitter,

I think he's funny, basically he says that if you fail at Kaggle competitions at first, you're either going to have to start improving your skills or you can go with Kaggle bashing. So, go ahead and do one of the two. And moreover, there are quite a number of real- world benefits for Kaggle competitions. So, remember I said that these are called inducement prize competitions. Basically, what happens is that people are asking Kagglers to take a look at a problem that the people who pose the problem don't know how to solve.

So, for instance, Kagglers can diagnose Melanoma better than medical doctors. There are a number of other examples along this line. And then additionally, again with respect to getting a job, a frustrating thing that I've seen on the job market is sometimes recruiters want to hire somebody who's already working on a particular problem, and it can be like, "Well of course I haven't worked on that yet, you haven't hired me yet." But Kaggle is kind of your shortcut into this kind of thing. So, basically what it is, is that if you are already working on something, because you have worked on a Kaggle competition on it, then you can say to a recruiter, "Oh yeah! I know how to do that" and prove that you know how to do that.

So, I think this is an extremely valuable area, basically both for fun and for sort of, for your career opportunities. And so, what we're going to do is that our capstone project is actually, exactly, the black-box optimisation competition that was from NeurIPS 2020. This competition is of course now finished, but it's extremely important. It's extremely important for several reasons. So first off, this is a type of programming challenge that comes up often in machine learning. So, you're going to have to tune parameters fairly frequently, it's needed for good algorithm performance. So, since you're going to need it for your jobs anyway, might as well have a base code that does it.

Secondly, this is a competition that recurs very frequently. So, in preparing for this particular capstone project, I looked up and I figured out that there are effectively equivalent competitions that I could find in every year from 2009 to 2021. So, this is a competition that I really suspect is going to come up again and again. Finally, I like this competition because I competed. If you want to go look at that black-box optimisation website, it, we had the name Cog Imperial, and you could see where I landed. So, it's quite a lot of fun. The competition will be going on throughout the course.

It's closed to students who are on this course, but of course you can use these exact skills for future competitions. And then finally, the tradition of Kaggle and

other competitions is that winners are going to publish their solution strategies. We'll do the exact same thing. So, at the end of this course, what we're going to do is we're going to ask everybody who did particularly well, to please give a short presentation explaining how it is that you did it right. And then that's useful for moving the field forward and indeed that's how, the people who work on Kaggle move their field forward. So, really looking forward to it. Hope you enjoy it too.