

Chapter 6.1- Machine Learning is Curve Fitting

13 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=6AMN0VE6B14](https://www.youtube.com/watch?v=6AMN0VE6B14)

<https://www.youtube.com/watch?v=6aMn0Ve6B14>

SUMMARY

Machine learning fundamentally revolves around the concepts of curve fitting and optimization, where the goal is to map input data to output predictions by finding the best-fitting model parameters. This introductory discussion emphasizes the importance of defining objective functions and understanding the differences between supervised and unsupervised learning paradigms.

- Machine learning can be viewed as a sophisticated form of curve fitting, where models are adjusted to best fit given data.*
- The model's parameters, denoted as Θ , are optimized to minimize a defined loss function, which measures the fit quality.*
- Supervised learning uses labeled data, making it easier to train models, while unsupervised learning involves finding patterns in unlabeled data, presenting a more complex challenge.*
- The training process typically involves dividing data into training, validation, and test sets to evaluate model performance accurately.*
- The choice of the objective function is crucial, as it influences how well the model learns and generalizes to new data.*
- Modern machine learning architectures often automate the process of data splitting and cross-validation to enhance model evaluation.*
- Understanding the structure and interpretability of neural networks is essential as they become increasingly complex in their parameterization.*

we started a new chapter and this chapter is really going to be bringing together two concepts that are very simple in some ways Curve fitting that we've already talked about and optimization and so this forms the basis of machine learning in fact I like this title here for this which is machine learning is curve fitting and that is in some I will adhere to that philosophy as we build forward we're going to do a very simple introduction of this now and of

course we're going to get much more sophisticated in later chapters when we really dive down on using machine learning but in many ways the view here is it is just a curve fitting process in other words you give me the function you want to fit to the data I have to find the parameters of that data in order for it to do to do a proper job fitting so let's start off with this and start framing why I would have this philosophy here

and it's not just my philosophy I think there's many people who believe this as well so machine learning is curve fitting remember that what we're trying to do in general in Curve fitting or even just modeling broadly is essentially take some data in and map it out to some data out so this could be a very general process so for

instance we could have data that we have that's X and what I'm trying to do is do a Best Line fit for

instance in other words how do I fit the best line to the data and now the output is a line from the data to a line or it could be I give you pictures of dogs and cats you tell me if it's a dog or a cat through a label or if I give you a a temporal sequence of data so here's the weather over the last month tell me the the weather tomorrow right so the input

could be essentially all these proxy variables for data like pressure temperature wind velocities you tell me what the weather is tomorrow so this is a very general structure for model building now okay so for an input to an output map now what you get to pick in this process then is this right here the F of θ and the F of θ f is the model you are imposing and θ are its

parameters so the model you're imposing this is why I call it a curvefit because essentially you as the expert in the loop are basically framing the problem are deciding what model you want to fit to this data for an outcome or an output and that model has a certain number of parameters and they're essentially characterized by θ now for see for deep learning and neural Nets in general θ is going to be quite large we're going to have a lot of different neural

network parameters and so that's going to be an important piece of this which is we have this idea of a very large high-dimensional space what we're trying to do then is find the variables θ that optimize this fit okay that's where it's like a curve fit let me just give you a simple example I could say I take some data I'm trying to do a line fit so here's my model my model is ax plus b

that is the line fit to some data my job when we do least Square fitting of a line is to find a and b so that cons that's my parameter say A and B I don't know what they should be I should try to pick a and b to do the best job in fitting the data so right away now we also have to Define what do we mean by the best job in fitting the data so let's come to

this and start thinking about optimization so we've talked about curve fitting and curve fitting we almost always just assumed we didn't really talk about an objective in curve fitting except for do a root mean Square error fit but more broadly an objective function is what we're going to rely on here in this optimization procedure which is the opt you know the optimization is set up with generically dependent upon X and Y is depending on θ and the objective

function could have lots of different pieces to it this is L_0 L_1 and the λ do determine how important some of the aspects of this objective function are so in other words I'm not just constrained to fitting through root mean square error I could constrain this fit to also have many other characteristics that I can express mathematically in L_1 through through L_{∞} for instance so this is important it gives you a very flexible framework in order for you to

decide what you want this model to give you okay and normally engineering this loss function or the objective function is an absolutely critical component for machine learning and deep learning in general in fact many of the best models out there this is where all the magic happens is really engineering this loss function so never lose sight of that you coming up with great L1 L2 L Zero and so forth and setting up the architecture of what you want your model to look

for in the first place is really critical okay so then once you have that loss function in other words the objective you would actually pick your model the F of θ you would essentially try to say I take my data in I have an output y Tilda okay and then I can actually ask the question so if I start off with some random parameters in θ I get some output and what I'm trying to do is minimize the loss

function so for instance here if the loss function is a root mean square error I look at what I get out from this model versus the truth of what the actual output is and I try to minimize that how do I minimize it I look for θ s you know I try to improve the value of θ s and this is going to be generically done through gradient descent type architectures okay and we'll talk quite a bit more about that

in later chapters but for right now mostly just is trying to build some intuition that this whole this entire infrastructure really is just about you coming up with a curve which can be very fancy in right here this F of θ but then we're just doing optimizing after that to find the variables of your curve okay so first let's talk about some of the different kind of AI algorithms out there for doing this or

at least that's some of the kind of different machine learning architectures that are available to us I'm going to start off with a couple here we also talk about reinforcement learning later in the book but supervised and unsupervised are the two dominant paradigms of learning supervised learning is much easier much more successful in general than unsupervised learning and it's just like they sound supervise means I actually have expert knowledge of the data and so I can label

the training data unsupervised is just I give you data your job is to find patterns and label it so the way you could actually construct this is first we need to think about the domain of the data we collect so typically we have some data sets that are in some n -dimensional space okay and that those data sets are a subset of all possibilities so D Prime we think about as all possible data D is the data I

have to build a model so there's usually a gap between D and D Prime and this Gap is really important in practice in a lot of machine learning but the fact is you typically have a finite number finite amount of data and that you're going to use to train your models on and but the real world is all the data in all possibilities which is D Prime okay and so now let's start setting up what it means to do

unsupervised learning unsupervised learning is I'm going to give you data and the inputs are going to be some examples X of J so this is going to constitute the input data Matrix and I have M data points and they're all n dimensional your job is to produce labels on the output so in other words you have data and your goal is to construct labels for the data so for instance if I gave you pictures of dogs

and cats but I didn't tell you they were dogs and cats I just put that into the computer could this algorithm actually start to identify which pictures are dogs which pictures are cats your job is to tell how many labels there are and in this so so and again you may not even know are there dogs and cats in here maybe there's kangaroos so these are the kind of things that you are faced with in modeling these systems

that are really important for you to consider so that's the basic structure of unsupervised learning you don't even know what these labels are and so you're trying to it's it's a much more difficult problem than if I actually knew what I was trying to find and supervised learning does know what it's trying to find other words when I give you the data I also give you labels so the data has labels with them so as instance I give you cat and dog pictures

and I tell you these are cats these are dogs this makes it much easier because I'm training this with expert knowledge and so now when I give you new data you just have to produce the labels out so now I give you the data you produce the labels but I've actually given the machine learning algorithm the ability to know what it's actually even trying to find what kind of labels it's actually trying to find so these are

important aspects of the machine learning that we have to think about and these are the two dominant paradigms I'm going to go into great deal on both of these much later in the book but essentially you can see what we're setting up here is all of this is a curve fit towards an input output relationship and these input output relationships are you know I give you data you give me a label for instance and so I'm trying to find a curve that

best does a job of labeling this so can be very generic in terms of its architecture all right so there's one last piece here that we should talk about which is when you go to train it's important to ask the questions how do I know when I've done a good job so from the data and the out input output pairs okay so here they are here let's call them x of j y sub j generically they're input output or they could be

input and the output of the labels but there is this idea of a training and cross validating set and then a test set so out of all that data for you to see how well you've done with your curve fit what you're going to have to do is break up your data into three pieces there's the training data and the training data has two pieces the training data and part of the training data is withheld to use as cross validation and then there's

a test set you cannot use your test set in training the test set only gets used after training so you randomly pull out some of the data and once you've trained your model you ask the question how well does it do on this test set that was never used in the training process okay so those are typically the three pieces that you're going to have to break up your data into the nice thing about a lot of modern machine learning

architectures it does this for you automatically and automatically cross validates for you it's such an important process to do this that it's there and another way to say it is I have a curve fitting process I want to test my curve fit so you do you test your curve fit on some new data that wasn't part of that training to see how well it did there okay so we come down to it then so then

the training procedure and again this is or if you want to call it a training procedure or just simply doing the curve fit itself is going to be you're going to specify an objective function and it can be very sophisticated objective function actually and this is one of the nice things about modern Computing is you can make very sophisticated architectures you come up with a model that relates the input to some output data and then you ask how do I update

the parameters of my model you do it by minimizing let's say for instance here the root mean square error but of course I could minimize something else it's all specified up here and so or a combination of things and so I here's my output here's what it's supposed to be and what I'm going to do is basically figure out how do I update the values of my parameters in order to minimize this in other words make my curve fit the

best possible so this involves Curve fitting and optimization and I'm going to walk you through examples of how this actually works in practice on very simple examples to show you how this actually works so that is the precursor to things and again just advocating that machine learning is curve fitting it's just the curves we're going to start using are really kind of exceptionally complicated and very highly parameterized at times but once you get used to them you actually see that

there's actually quite a bit of structure quite a bit of interpretability in terms of building neural network architecture themselves