

Why human-level performance?

(DESCRIPTION)

A man addresses the camera. Text. Comparing to human-level performance. Why human-level performance?

(SPEECH)

In the last few years, a lot more machine learning teams have been talking about comparing the machine learning systems to human level performance.

Why is this?

I think there are two main reasons.

First is that because of advances in deep learning, machine learning algorithms are suddenly working much better and so it has become much more feasible in a lot of application areas for machine learning algorithms to actually become competitive with human-level performance.

Second, it turns out that the workflow of designing and building a machine learning system, the workflow is much more efficient when you're trying to do something that humans can also do.

So in those settings, it becomes natural to talk about comparing, or trying to mimic human-level performance.

Let's see a couple examples of what this means.

(DESCRIPTION)

Text. Comparing to human-level performance. An X-Y graph with vertical axis labeled accuracy and horizontal axis labeled time. A horizontal dotted line representing human is high on accuracy.

(SPEECH)

I've seen on a lot of machine learning tasks that as you work on a problem over time, so the x-axis, time, this could be many months or even many years over which some team or some research community is working on a problem.

Progress tends to be relatively rapid as you approach human level performance.

But then after a while, the algorithm surpasses human-level performance and then progress and accuracy actually slows down.

And maybe it keeps getting better but after surpassing human level performance it can still get better, but performance, the slope of how rapid the accuracy's going up, often that slows down.

And the hope is it achieves some theoretical optimum level of performance.

(DESCRIPTION)

A horizontal dotted line above the human line labeled Bayes optimal error.

(SPEECH)

And over time, as you keep training the algorithm, maybe bigger and bigger models on more and more data, the performance approaches but never surpasses some theoretical limit, which is called the Bayes optimal error.

So Bayes optimal error, think of this as the best possible error.

(DESCRIPTION)

Text. Best possible error

(SPEECH)

And that's just the way for any function mapping from x to y to surpass a certain level of accuracy.

So for example, for speech recognition, if x is audio clips, some audio is just so noisy it is impossible to tell what is in the correct transcription.

So the perfect error may not be 100%.

Or for cat recognition.

Maybe some images are so blurry, that it is just impossible for anyone or anything to tell whether or not there's a cat in that picture.

So, the perfect level of accuracy may not be 100%.

And Bayes optimal error, or Bayesian optimal error, or sometimes Bayes error for short, is the very best theoretical function for mapping from x to y .

That can never be surpassed.

(DESCRIPTION)

A sloped purple line represents work on the problem. It increases rapidly over time until it passes human performance, then levels out as it approaches but does not reach the Bayes optimal error.

(SPEECH)

So it should be no surprise that this purple line, no matter how many years you work on a problem you can never surpass Bayes error, Bayes optimal error.

And it turns out that progress is often quite fast until you surpass human level performance.

And it sometimes slows down after you surpass human level performance.

And I think there are two reasons for that, for why progress often slows down when you surpass human level performance.

One reason is that human level performance is for many tasks not that far from Bayes' optimal error.

People are very good at looking at images and telling if there's a cat or listening to audio and transcribing it.

So, by the time you surpass human level performance maybe there's not that much head room to still improve.

(DESCRIPTION)

Arrow indicates the small gap between human and Bayes.

(SPEECH)

But the second reason is that so long as your performance is worse than human level performance, then there are actually certain tools you could use to improve performance that are harder to use once you've surpassed human level performance.

So here's what I mean.

(DESCRIPTION)

Text. Why compare to human-level performance? Humans are quite good at a lot of tasks. So long as M-L is worse than humans, you can:

(SPEECH)

For tasks that humans are quite good at, and this includes looking at pictures and recognizing things, or listening to audio, or reading language, really natural data tasks humans tend to be very good at.

For tasks that humans are good at, so long as your machine learning algorithm is still worse than the human, you can get labeled data from humans.

That is you can ask people, ask higher humans, to label examples for you so that you can have more data to feed your learning algorithm.

Something we'll talk about next week is manual error analysis.

But so long as humans are still performing better than any other algorithm, you can ask people to look at examples that your algorithm's getting wrong, and try to gain insight in terms of why a person got it right but the algorithm got it wrong.

And we'll see next week that this helps improve your algorithm's performance.

And you can also get a better analysis of bias and variance which we'll talk about in a little bit.

But so long as your algorithm is still doing worse than humans you have these important tactics for improving your algorithm.

Whereas once your algorithm is doing better than humans, then these three tactics are harder to apply.

So, this is maybe another reason why comparing to human level performance is helpful, especially on tasks that humans do well.

And why machine learning algorithms tend to be really good at trying to replicate tasks that people can do and kind of catch up and maybe slightly surpass human level performance.

In particular, even though you know what is bias and what is variance it turns out that knowing how well humans can do on a task can help you understand better how much you should try to reduce bias and how much you should try to reduce variance.

I want to show you an example of this in the next video.