Summary

Breiman compares traditional statistics (assuming the process the data was generated in a parametric approach) with machine learning, which typically does not make any such assumptions. He says that making such assumptions is dangerous as they could very easily be misspecified. He shows that ML typically gives better accuracy, which he argues is the most important part of statistical analysis.

Reaction

Very interesting paper – I liked how he conveyed his information, he didn’t talk super formal or academic and I feel like it’s easier to follow this way. It is always interesting to see what people were doing and thinking when currently prominent fields (like ML) were up and coming.

Questions

1. One thing Breiman says is that these models should be used for inference as well, since the better your model is the more useful information you can get out of it. As far as I know, this is very much not the case today, I’ve never seen anyone doing inference that is concerned about model performance. So, I’d be curious to know if that was ever considered or what the inference people think about it.
2. Also, is the above point moot? I’m wondering if inference people will ever note the accuracy of the model they use and note if the inference can be relied upon if the accuracy is bad. I don’t think this happens, but just checking