Q9: Is the Rashomon set realistic?

Rashomon set considers that if the data allows a large set of reasonably accurate predictive model to exist, we are bound to have at least one interpretable model, thus rendering it accurate and predictable (Rudin, 2019).

Based on this logic, models from Rashomon set are better than explainable models to a certain extent. As black box machine learning models are not only uninterpretable, they are also complicated and difficult for human to make sense of it(Rudin, 2019). Thus, Rashomon set offers an alternate view for people as opposed to accepting the ambiguity of black-box machine learning model.

Should the black box machine learning models be more transparent in the first place, Rashomon set would not necessary be used to capture explainable models (Ferrini, 2020). Building a transparent and interpretable data right from the start seems more logical, rather than building models and explaining them later with the possibility of it not even making sense.

However, interpretable models may not make sense too. The title of paper states: ‘Stop explaining black box machine learning model and use interpretable model instead’. We cannot say for sure that interpretable model truly capture explainable model? It may not be better than explainable model.

Rudin has stated the difference between explainable and interpretable models. Explainable may not mean that the model is interpretable (Rudin, 2019). Even though large majority of people prefers interpretable models, it can also be difficult to prove that the explainable models are not interpretable.

The seemingly less popular black box machine learning explanations may not have enough details for human to fully understand and it could just be approximated models. Contradictorily, are not all models approximated too? Even if the model is interpretable, the model itself is also an approximation.

Also, the Rashomon set needs to have a define domain-specific method before it can be searched for an interpretable model. The belief in the trade-off between accuracy and interpretability deters developers to create interpretable models with the excuse of it being less accurate (Martin, 2019).

In these cases, Rashomon set is not realistic in meaningfully capture explainable models.

Quoting George Box, “All models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.” (George, 2007). Research is endless and there will always be room for improvement. What we can do is with each step we take, we try to be less wrong than the previous one.

Ultimately, our goal from building these models is to tell our audience a story with the data we have. How we make sense of the data, mainly depends on our what we have learnt and how we perceive things. Just like many things in life, data and models are just numbers and symbols until we give it a meaning. How successful a model is, is also dependent on how agreeable majority of the people are.

Based on the origin of Rashomon effect, what a person sees depends on where the person stands. People all look at things from different perspective. Just like a Rashomon set has many different models, each of them similar yet different, every individual may also interpret each single model differently. The possibilities are endless.

No matter how much sense we try to make out of these models, it may still be senseless. Or in a more optimistic perspective, some models may make more sense than the others. There is no best or true model out there. Rashomon set is only as realistic as we deem it to be.

References

1. Ferrini, M. (2020, January 06). Shall we build transparent models right away? Retrieved February 24, 2021, from <https://towardsdatascience.com/shall-we-build-transparent-models-right-away-196db0eeba6c>
2. George E. P. Box, Norman R. Draper (2007). “Response Surfaces, Mixtures, and Ridge Analyses”, p.414, John Wiley & Sons
3. Martin, T. (2019, August). Retrieved February 24, 2021, from <https://www.mlmi.eng.cam.ac.uk/files/tam_final_reduced.pdf>
4. Rudin, C., Wang, C., & Coker, B. (2019, June 17). The age of secrecy and unfairness in recidivism prediction. Retrieved February 24, 2021, from <https://arxiv.org/abs/1811.00731>
5. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. (n.d.). Retrieved February 24, 2021, from <https://www.kdnuggets.com/2019/11/stop-explaining-black-box-models.html>