

## **Question 9**

The Rashomon set, in theory, sounds realistic. For any given data, there exists a corresponding set of reasonably accurate models that explains it, and we can expect at least one of these models to be interpretable. The claim that accurate and explainable models exist in place of complex ones is irrefutable, but whether they can be meaningfully captured is a different question altogether. In practice, the Rashomon set hardly comes into play, making it idealistic.

The author herself has raised key points on why it is challenging to create interpretable models, and they can be summarised by one line in the paper – *Solving constrained problems is generally harder than solving unconstrained problem*. The goal of every model is to mimic the underlying relationship between the response and the predictors. Of course, the true relationship is something we will never uncover, and we can only aim to get as close to it as possible.

The problem with finding interpretable models is that in trying to make these models more explainable, we end up making more simplifying assumptions on the underlying relationship of our data. This makes it even harder for our models to get closer to the truth. Adding constraints limits the type of models that we can find.

This affects the way we construct the Rashomon set, and whether we can make it as inclusive as possible. To be able to meaningfully capture explainable models, we need to ensure first that our set is comprehensive. This is where we find ourselves in a catch-22 – for our Rashomon set, we need to find explainable models, but simplifying the model makes it harder to find those that do fit in the set.

It is much easier to create black box models that fit in the Rashomon set rather than interpretable models that fit the set. The author does make valid points, there must exist accurate, explainable models for any given dataset, but she underestimates how hard it is exactly to identify these models.

That is also not to say that we should not try to find these models. The proposed solution of mandating organizations to report the accuracy of interpretable models is a valid one. Explainable models should always be used in place of black box models, and despite it being harder to create simple, accurate models, they are an avenue that should still be explored.

While in theory, the Rashomon set is sound, it is not that easy and implementable. We know that explainable models exist, but the challenge lies in capturing them. It is more accurate to call the Rashomon set idealistic, rather than realistic.