Question 9’s Answer

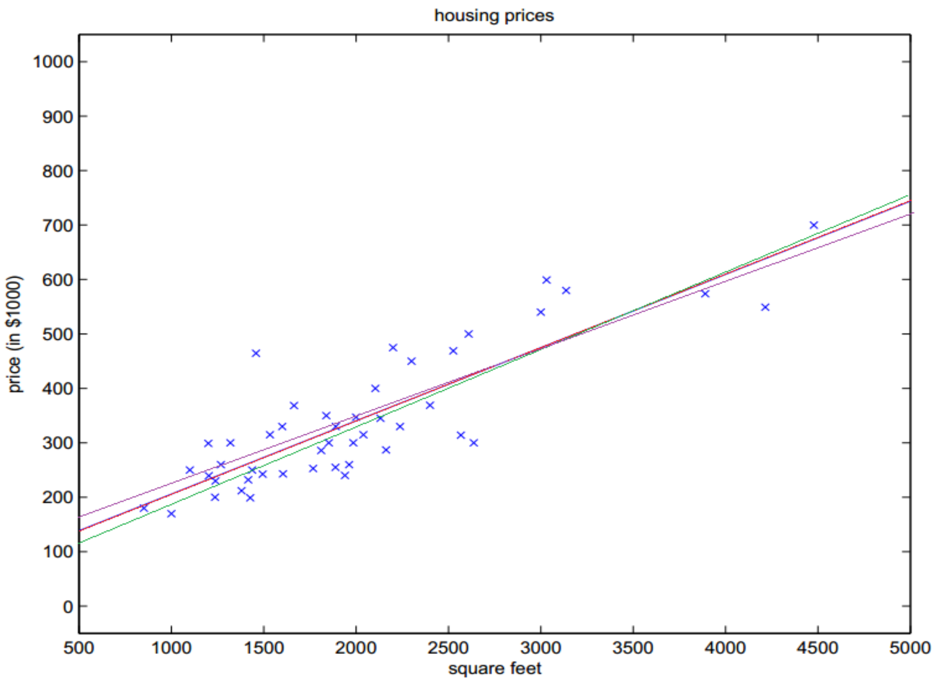
I think the Rashomon set is realistic.

In the article provided by the title alone, the authors cite a large number of papers (among which interpretable models are cited) to illustrate the progress of interpretable work and specific examples of successful applications in specific fields (e.g., physics and medicine), which suggests that interpretable models are feasible in at least some specific fields.

For a given dataset, using the black box model as an example, we can use different network structures, different optimization methods and tuning techniques to obtain different network models, and the resulting models may be similar in terms of prediction accuracy.

In the feature dimension of data set, there may be some redundancy or correlation. If we can sort out the redundancy, correlation and causality in data preprocessing, we can structure the data and establish an interpretable model.

Taking the linear regression in low dimensional space as an example, we can construct multiple interpretable models with different methods in a limited data set. The figure below describes the relationship between house price and room area. There may be multiple interpretable models. But which one is the best?



The specific house price may also be related to the number of rooms in the house, which needs to add more appropriate constraints when building the interpretability model.

For deterministic, complete and clean data, it is much easier to use black-box machine learning methods than to troubleshoot and solve computationally difficult problems. However, for high-stakes decisions, high-latitude and messy data, it requires expertise in building interpretable models or a more specialized knowledge base in the domain. And obtaining interpretable models usually requires establishing such a specific set of constraints that solving constrained problems is usually more difficult than solving unconstrained problems, which is a very challenging task.

In a specific field, we sort out some structured information from the limited data set, which can build interpretable models. Because the available data is limited, we have some uncertainty in the limited data. Data can hold many models close to the optimal solution, which may use different methods and functions, but their results can be equivalent. For some samples that do not meet the expectation, more energy and professional knowledge may be needed to find out all the features that affect our desired results and the relationship between them, and to expand and improve the structured information to narrow the Rashomon set, so as to approach the real solution.

Taking image processing as an example, we can obtain a large number of image data of the same scene under certain conditions. For this scene, we can get one or more groups of good threshold parameters through debugging and experiments. Through this group or groups of thresholds, we can obtain the image information we care about. These group thresholds and processing functions can be regarded as Rashomon sets. However, for other scenes, even for the same scene with different weather or lighting conditions, the processing results are not as good as they could be. This indicates that, for the uncertainty of the generated data, even if we consider each pixel dimension, we still cannot approximate the optimal solution for all scenes, which may require us to add more constraints, such as the association between pixel values, image content contours, angles, binarized image analysis, histogram distribution, and other possible additional considerations when building interpretable models, in order to constrain and adjust the already existing Rashomon set and thus capture the interpretable model.

Therefore, I believe that Rashomon Set is meaningful for capturing interpretable models. The existence of multiple solutions is not conflicting, either in terms of dealing with practical problems, or mathematically. For the interpretable model in Rashomon set, it is necessary to further expand and adjust the constraints to adjust the interpretable model when it is unable to adapt or approach the real solution. It is meaningful in the construction and generation of the whole interpretable model.