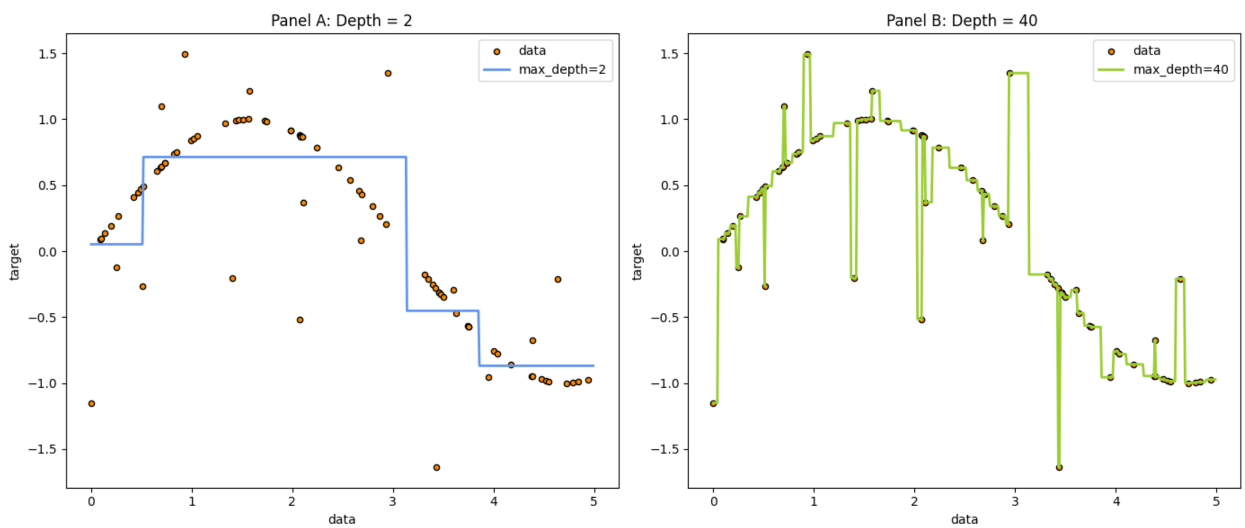
<u>**A Note on Model Complexity, Bias and Variance**</u>

We progressively moved towards more complex models in this course. Do more complex models necessarily mean better out-of-sample prediction? This note addresses this issue.[1]

- **Bias-Variance Tradeoff**

  For context, consider the example in 'Regression Tree Example.ipynb.' Panel A fits in the chart a regression the tree with max_depth = 2 and Panel B fits a regression tree with max_depth = 40.



- **Bias**: Bias quantifies the systematic error introduced by approximating a complex real-world function with a simplified model based on what we can learn from a finite training dataset. For example, in the scatter plot above, the data were generated from an underlying sine function with added noise, but we observe only a particular training data set. Panel A approximates the function using a regression tree of depth 2, which results in a high bias because the model is too simple to capture the underlying structure of the training data. In contrast, Panel B, with a more complex regression tree, has a lower bias because it better fits the training data.
  - **Key Takeaway:** The best fit in a specific training sample minimizes error for **that sample**, but it may overfit noise rather than capture the general pattern and hence may generalize poorly.

- **Variance**: Variance quantifies the sensitivity of a model's predictions to fluctuations in the training dataset. It measures how much the model's predictions vary for a given test dataset when trained on different samples. Since different training datasets contain

[1] See James et al. (Chapter 2) for a detailed discussion of Bias-Variance tradeoff.

different realizations of the noise term, a more complex model, such as a regression tree with depth = 40, is highly flexible and captures more of the noise in the training data, leading to high variance. In contrast, a simpler model with depth = 2 exhibits lower variance because it generalizes better across different datasets.

- **Bias-variance tradeoff:** A model that generalizes well should balance bias and variance across different training datasets, rather than optimizing only for a single dataset. Heuristically, the total prediction error on test data can be decomposed into bias and variance components. As the figure below shows, total error initially decreases with increasing model complexity, as the model captures more of the underlying pattern. However, beyond a certain point, increasing complexity leads to overfitting, where the model starts capturing noise rather than the true signal, resulting in higher variance and increased overall error.