Talking Points

Slide 10

When building a machine learning model to predict house prices based on features like size, location, and number of bedrooms, grid search can help systematically explore different combinations of hyperparameters. By tuning these hyperparameters, models can be optimized to make more accurate predictions.

In the decision tree model, each node represents a decision point based on features like location or size. The depth of the tree determines how many layers of decisions the model can make. A deeper tree allows for more complex decision-making as it can consider more features and their interactions.

Think of a decision tree model like a game of 20 Questions. The model asks questions about a house's features to guess its price.

Now, hyperparameters are settings that make the game work better. For example:

The maximum depth of the tree is like saying how many questions the model can ask before making a guess. If it's too deep, the game might get too complicated.
The minimum number of samples to split a node is like deciding how many houses need to be in a group before the model can ask a question about them.
So, in simpler terms, hyperparameters are just rules that make the game of 20 Questions (or the model) play better.

Finding the right balance between depth and the number of nodes is crucial. Too shallow a tree might oversimplify the model and miss important patterns, while too deep a tree could lead to overfitting, where the model memorizes the training data instead of learning general patterns.

Overfitting is a common issue in machine learning, occurring when the model fits the training data too closely, capturing noise instead of underlying patterns. In a regression model, especially when working with many features or a limited dataset, using ridge regression can reduce the opportunity for overfitting by introducing a penalty term. This penalty or cost discourages the model from relying too heavily on any one single feature, promoting a more balanced model fit.

NEXT SLIDE

Slide 11

By applying these different models and optimization methods, we were able to gain insights on how optimization methods influence accuracy.

Our analysis reveals that different optimization methods lead to varying levels of model accuracy. For instance, while linear regression with ridge regression achieves an R-squared value of 0.6233, decision tree and random forest models yield higher R-squared values of 0.6942 and 0.6711, respectively. This highlights the significance of selecting the appropriate optimization method tailored to the dataset and problem at hand.

CONCLUSION SLIDE