Hyperparameter Optimization: Takeaways

Syntax

- Using the expand.grid() function to automatically create hyperparameter combinations:

```
knn_grid <- expand.grid(k = 1:20)
```

- Plotting to visualize the optimal k value:

```
knn_model <- train(tidy_price ~ accommodates + bathrooms + bedrooms,
                        data = training_data,
                        method = "knn",
                        trControl = train_control,
                        preProcess = c("center", "scale"),
                        tuneGrid = knn_grid)
    plot(knn_model)
```

Concepts

- **Hyperparameters** are parameters that affect the behavior and performance of a model, but are unrelated to the dataset itself.
- **Hyperparameter optimization** is the process of finding the optimal hyperparameter value, given a dataset and a machine learning model
- **Grid search** is a simple but common hyperparameter optimization technique, which involves evaluating the model performance at different `k` values and selecting the `k` value that results in the lowest validation error. Grid search involves:
    - Selecting a subset of the possible hyperparameter values.
    - Training a model using each of these hyperparameter values.
    - Evaluating each model's performance.
    - Selecting the hyperparameter value that resulted in the lowest error value.

- The general workflow for finding the best model is:
    - Selecting relevant features to use for predicting the target column.
    - Using grid search to find the optimal hyperparameter value for the selected features.
    - Evaluate the model's accuracy and repeat the process.

Resources

- Difference Between Parameter and Hyperparameter
- Hyperparameter Optimization
- caret Documentation