

Bagging & Pasting

Bagging (Bootstrap Aggregating)

Bagging involves creating multiple subsets of the original training data through random sampling with replacement (bootstrap sampling). Each subset is used to train a separate model, typically of the same type (e.g., decision trees). The final prediction is made by aggregating the predictions of all individual models, usually through voting for classification or averaging for regression.

Key points of bagging

1. Sampling is done with replacement, so some instances may appear multiple times in a subset.
2. Each model is trained independently.
3. Reduces variance and helps prevent overfitting.
4. Random Forest is a popular example of bagging with decision trees.

Pasting

Pasting is similar to bagging but uses sampling without replacement. This means each data point can only appear once in each subset.

Key points of pasting

1. Sampling is done without replacement.
2. Like bagging, it creates multiple subsets and trains independent models.
3. Can be useful when dealing with very large datasets.

Bootstrap Sampling

- Bagging creates multiple subsets of the original training data through random sampling with replacement.
- Each subset, called a bootstrap sample, has the same size as the original dataset.

Out-of-Bag (OOB) Estimation

- Samples not used in a particular bootstrap can be used to estimate the model's performance without a separate validation set.

