

Predictive modeling and machine learning

Kohei Arai

1

7. Methods for future predictions and advanced data analysis

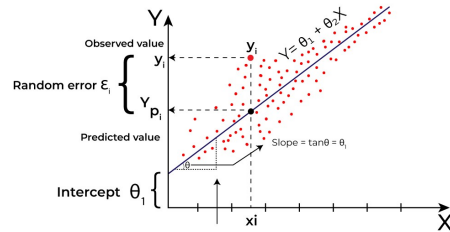
- Linear Regression: A prediction model using simple linear relationships
- Kernel Ridge Regression & Lasso regression: A regularized regression method that can also handle nonlinear relationships
- Gradient Boosting: A method that combines multiple weak learning models to perform highly accurate predictions
- Visitor Prediction: Predicts future numbers of visitors from past data
- Churn Prediction: Predicts future churn from past churn data and market trends

2

Linear regression

- A basic example of linear regression using the scikit-learn library in Python.
- This example shows both simple regression (when using one explanatory variable) and multiple regression (when using multiple explanatory variables).
- Use the LinearRegression class to create a linear regression model and fit it to the training data.
- Finally, use the trained model to make predictions on new data and visualize the results.

3



4

Nonlinear regression

- As an example of nonlinear regression, we will use the scikit-learn library to create Kernel Ridge Regression.
- Kernel Ridge Regression (as well as Lasso regression) is suitable for capturing nonlinear relationships.
- We used synthetic data with noise added to the sine function.
- We used the KernelRidge class to create a kernel ridge regression model and specified the RBF (Radial Basis Function) kernel.
- Finally, we visualized the training data and the regression curve.
- There are various other nonlinear regression methods, but the method you choose will depend on the specific problem.
- To choose the appropriate method, it is important to consider the method that suits the characteristics of the data and the purpose.

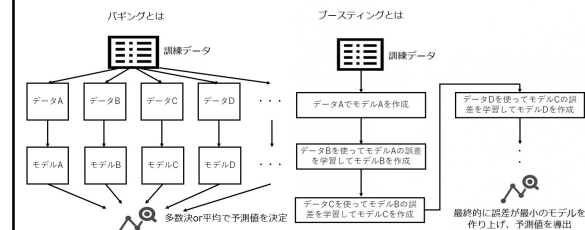
5

Gradient Boosting

- As code examples for Gradient Boosting, we present GradientBoostingRegressor (regression) and GradientBoostingClassifier (classification) using the Python scikit-learn library.
- Use GradientBoostingRegressor and GradientBoostingClassifier from scikit-learn.
- Load the appropriate dataset and set the objective and feature variables appropriately: hyperparameters (e.g. n_estimators, learning_rate, max_depth) need to be tuned depending on the specific problem.

6

RandomForest : Bagging LightGBM, Xgboost : Boosting



7

Customer forecasting

- To show the code for customer forecasting, we need specific requirements, such as what kind of data you want to use to make the prediction, the type of prediction model, and the use of specific libraries.
 - We will show an example of predicting future customer numbers using simple time series data.
 - We will use a time series forecasting library called Prophet.
- Install Prophet.
pip install fbprophet

8

Customer forecasting

- Convert time series data into a format suitable for Prophet and train the Prophet model to predict future visitor numbers.
- Data preprocessing and model adjustment are required to match the actual data and requirements.
- There are various other methods for visitor forecasting, and the appropriate method varies depending on the specific scenario → It is important to select a model according to the nature of the data and the forecast period.

9

Churn forecasts

- Churn forecasts are made using various methods and models.
- Predict churn using time series data: Use a time series forecasting library called Prophet.
- Convert time series data into a format suitable for Prophet and train the Prophet model to predict future churn.
- Data preprocessing and model adjustment are required to suit the actual data and requirements.
- There are various other methods for churn forecasting, and the appropriate method varies depending on the specific scenario → It is important to select a model according to the nature of the data and the forecast period.