



Exercises

Interpretation Challenges I

- 1 What is the difference between Regression and Classification?
- 2 Read Flight data
- 3 Print all features
- 4 The goal is to reduce the cost of flight delay.
Which target feature do we choose and why?
- 5 How can we determine how strong is the influence of WEATHER_DELAY on ARR_DELAY?
- 6 Delete ARR_DELAY and String data
- 7 Set the target of a delay >15 minutes to 1 otherwise to 0

- 8 How linear correlated are the features to the target?
- 9 What are the requirements of Linear Regression?
- 10 Test Normal Distribution (with Kolmogorov-Smirnov test) of the feature `ARR_DELAY` (use only the 100 data sets) and print the histogram
- 11 Perform a nonlinear transformation of the values and check again for normal distribution
- 12 Use a decision tree for prediction.
- 13 Print mean squared error and R2 score.

14 To compare different models create a dictionary of the models

- Linear Regression
- Decision Tree
- MLP (MLPRegressor)

15 Fit all these models and print RMSE train, RMSE test and R2 score for test data as HTML table

16 Use for the classification of the flight delay Logistic Regression, Decision Tree and Gradient Boosting

17 Compare the classification methods using AUC, Recall, F1 score

18 Print the ROC for all models

19 Print the Confusion Matrices for all models