

Exercise Sheet 3

Due date May 26, 2025

1 Task: Rule Learning

- Install the `imodels` library for rule learning, if you have not done so already¹.
- Load the bike rental dataset similar to the previous exercise sheet.
- Rule learners can typically only handle classification problems but not regression problems. Hence, convert the regression problem into a classification problem by discretizing the label (number of rented bikes), e.g., by specifying whether the number of rented bikes is above or below a certain value (e.g., the median).
- Similarly, rule learners typically require categorical features (instead of continuous features). Convert the continuous features to discrete features, for example, via the Pandas method `qcut` and then use dummy encoding.
- Learn rules via the `GreedyRuleListClassifier` of `imodels`. What is the training accuracy and the test accuracy of the classifier? How can the rules be understood? For example, what does “risk” mean? What does “pts” mean? How are “risk” and “pts” related to the “coverage” and “accuracy” of a rule as introduced in the lecture?

2 Task: PDP, ICE, Permutation Feature Importance

- Use the bike rental dataset from exercise sheet 2, fit a random forest² and plot the partial dependence plots (PDP) for the features Temperature, Humidity, and Windspeed with `scikit-learn`.³
- Compute the permutation feature importance from scratch (just by using `numpy` or `pandas`). After that, use the `permutation_importance`⁴ function from `scikit-learn` to make sure that your implementation is correct. Finally, plot the results with error bars⁵.

¹If you receive the error “ImportError: cannot import name ‘indices_to_mask’”, you need to update `scikit-learn` to version 1.5

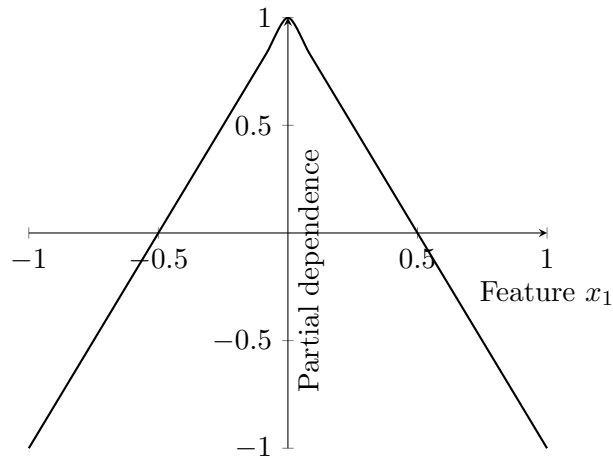
²<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

³https://scikit-learn.org/stable/modules/partial_dependence.html

⁴https://scikit-learn.org/stable/modules/permutation_importance.html

⁵https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.errorbar.html

- c. Create a dataset with two features and use a RandomForestRegressor such that the PDP of the first feature has approximately the following form: The plot goes from -1 to 1 both on the x and the y axis. For negative x values, it has a positive slope and for positive x values, it has a negative slope (see figure below). Briefly explain how you chose the feature and target values and why this leads to such a plot.



- d. Create a dataset with two features such the PDP of the first feature is (approximately) constant but the permutation importance of the first feature is high. Briefly explain how you chose the feature and target values and why this leads to such a plot.
- e. Create a dataset such that the PDP is (approximately) constant but the individual condition expectations (ICE) diverge (e.g., one goes up and one goes down). Briefly explain how you chose the feature and target values and why this leads to such a plot.