VU Machine Learning

Wintersemester 2022

Exercise 2

Nysret Musliu (nysret.musliu@tuwien.ac.at)

Exercise 2

- Groups of 3 students
- Implement techniques for classification/regression
- Compare to existing techniques
- Submit the source code
- Prepare around 10-20 slides
- Individual discussions (30 min) for each group (all members must be present)
- Submission: December 12
- Presentations: December 13-16, January 09-11 (slots will be available in tuwel)

Exercise 2 – Techniques

- Implement a Naïve Bayes algorithm
  - Your algorithm should work for classification data sets that include nominal or/and numeric attributes (features)
- Implement one of the algorithms considered in the class for regression (you can select the algorithm)
  - Implement also a simple method to automatically find the best values for the hyperparameters
- You should implement these algorithms from scratch
- Please do not use any part of existing code
- You can use existing code/functions for general parts like
  - Partial derivatives, cost function, distance functions…
  - Code for reading the input and testing the algorithm (cross-validation, performance metrics…)

Exercise 2 – Data Sets

- Pick 2 classification and 2 regression data sets
  - Data sets from the previous assignments can be used
  - Other data sets from UCI ML Repository, Kaggle…
- Should have different characteristics
  - number of samples – small vs. large
  - number of dimensions – low vs. high dimensional
- Pre-process the data set if needed (scaling, missing values …)

Comparison

- Compare your implemented techniques with
  - Existing implementations of techniques that you implemented
    - You can use the default parameters for the existing techniques
  - + Another classification/regression technique
- Apply cross-validation
- Conclusions
  - How efficient are your algorithms
  - Performance of your algorithms regarding performance metrics for classification/regression
  - Impact of hyperparameters
  - Other findings

Submission

A zip file with

- **Source code:**
  - You can use any programming language: Python, Mathlab, R…
  - Provide the information for the packages needed to run you code
- **Slides**
  - Around 10- 20 slides
  - No report needed
- Submission deadline: December 12, 22:00

Slides

- A brief discussion of implementations
- Comparison with the existing implementations/other algorithms
- Discussion of experimental results
- Conclusions/lessons learned

Individual discussions

- Length of discussion: 30 minutes
- Questions about
  - Source code
  - Techniques
  - Comparison with the existing techniques
- All members of the group should be able to explain the code/experiments
- The evaluation will be based on your code, discussion, comparison, and conclusions/lessons learned