

# VU Machine Learning

Summer Semester 2024

## Exercise 2

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

- Groups of 3 students
  - Implement machine learning techniques for predicting numeric values
  - Compare to existing techniques
  - Submit the source code
  - Prepare approximately 15-30 slides. No report is required for this assignment
  - Individual discussions for each group (all members must be present)
  - Submission: May 27th (morning)
  - Discussions: after May 26th (slots will be available in Tuwel)
- .....

- Implement the **gradient descent algorithm** for linear regression
  - You can refer to the algorithm provided in the lecture slides and use the same cost function
  - Experiment with different learning rates
- Implement the **k-nn algorithm** for regression
  - Experiments with different k and distance functions
- 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 for k-nn
  - Code for reading the input and testing the algorithm (cross-validation, performance metrics for regression...)

- Compare your implemented techniques with
  - The existing implementations of gradient descent and k-nn
  - One other regression techniques (e.g., regression trees, random forest,...)
    - You can use the default parameters for the existing techniques
- Use at least two performance metrics for comparison
- Conclusions
  - How efficient are your algorithms
  - Performance of your algorithms regarding performance metrics for regression
  - Impact of learning rate, k, distance functions
  - Other findings

- Pick 3 regression datasets 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, ...)

A zip file with

- **Source code:**
  - You can use any programming language: Python, R, Matlab, ...
  - Provide the information for the packages needed to run your code
- **Slides**
  - Around 15 - 30 slides
  - No report needed
- Submission deadline: May 27th, 9:00 a.m

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

- Length of discussion: 30 minutes
- You will have questions about
  - Implementation/Source code
  - Theoretical questions about techniques
  - Comparison with the existing techniques
- All members of the group should be able to explain the code/experiments
  - Students in the same group can receive different numbers of points based on the discussion
- The evaluation will be based on your code, discussion, comparison, and conclusions/lessons learned