Neural and Evolutionary Learning

Final Project Guidelines

Overview

The primary objective of the project is to compare and discuss the behaviour, performance, and application of the models studied in the NEL course.

For this purpose, the dataset located in the Practical Class 02/datamart folder should be utilised. This dataset has data on cow milk production from a farm in North Italy, which is entirely based on Automatic Milking Systems (AMSs). In the AMS adopted in this farm, cows decide when they will be milked. During each milking event, the milking robot acquires extensive data on the cow level, productivity and milking behaviour.

In particular, milk contents (fat, protein, lactose) are measured at each milking event and are used to evaluate the milk quality. The measurement of these components by the milking robot is not straightforward: it is done with a colourimetric method, and the robot needs to be calibrated every 2 weeks with chemical measurements from the laboratory.

Modelling problem

The modelling problem to be solved in the project is to infer the lactose contents exclusively using data from the milking robots.

The features dataset is provided in the data_project_nel.csv file. Lactose target values are provided in the y_lactose.csv file. *Additionally*, students can also work with models for Fat (y_fat.csv) and Protein (y_protein.csv) contents.

For a detailed description of the features, check the file Practical Class 02/nel_data_dictionary.pdf on Moodle.

Notice that, as shown in the data exploration notebook, the dry_days feature has 147 missing values. They are all observations of the first lactation period. This is simply because before the first lactation cows do not have a dry period. Therefore, it doesn't make sense to input data, and students should choose one of the following approaches:

- To model only data of the first lactation, not using the dry_days feature;
- To model only data of the second lactation onwards, using the dry_days feature;
- To model data of all lactation periods, not using the dry_days feature;

Models to be compared

At least four models should be compared:

- Genetic Programming (GP)
- Geometric Semantic Geometric Programming (GSGP)
- Neural Networks with backpropagation (NN)
- NeuroEvolution of Augmenting Topologies (NEAT)

The comparison among the algorithms should be done using statistical analysis, as indicated in Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. Sci Rep 14, 6086 (2024).

Groups

The project should be done in groups of 3 or 4 people.

If the student presents a research proposal with a feasible deliverable to be concluded before the project deadline, it is not necessary to have a group.

On the other side, if the student works on the proposed project, he/she must have a group. Otherwise, there will be a penalty, as working in a team is a very important skill.

Deliverables

A written report and the codes should be sent by 23h59 on June 9th to krebuli@novaims.unl.pt. The report should be no longer than 4 pages, including plots and references. The focus should be on discussing the results and the models' comparison. The codes should be done in Python, and plug-and-play for a new dataset.

Oral evaluation

During oral evaluation, all students will answer at least one question regarding the comparison and/or behaviour of the studied algorithms. Additionally, there will be at least 2 questions, one regarding the code implementation and another regarding the implementation of the algorithms. The students can be selected to answer these questions directly by the Professor or by chance, according to the group division.

Evaluation Criteria

The table below shows the evaluation criteria of the project.

Category	Subcategory	Points	Total
Source code	Code organization	1	8
	Code efficiency	1	
	Classes basic implementation (to be informed in each class)	2	
	Classes advanced implementation (to be informed in each class)	4	
	Extra implementation	*	
Report	Writting (maximum 4 pages, including references)	1	4
	Plots	1	
	Statistical methods	1	
	Discussion of results	1	
	Extra page **	-2	
Oral presentation	Working group distribution	1	8
	Code implementation discussion	2	
	Algorithms implementation discussion	3	
	Individual alogorithms comparison discussion	2	
		Final grade	**20**

Obs: * It depends on the complexity of the implementation.

** Maximum of 2 extra pages. Reports longer than 6 pages will not be accepted.