Training, Validation and Testing

Training Data

Three years of hydroelectric dam energy price data is given in the file ***train.xls***. These three years should be used to train your algorithm (Tabular Q-Learning, Deep Q-Learning etc.) on.

Validation Data

An additional two years of hydroelectric dam energy price data is given in the file ***validate.xls***. These two years should be used to test the algorithm you trained on the ***train.xls*** data. Using this validation set will give you a rough estimate on how well your algorithm responds to unseen data, and thus how well it could perform on the final graded test set.

Testing Data

Finally, at the end of your project, we will test your algorithm on the test dataset called **test.xls**, which is of exactly the same size as ***validate.xls*** (2 years of price data). The performance of your algorithm on this test set will influence the grade of your project. This dataset will therefore not be available to you during the project.

Code Requirements

It is important to set up your code so that it can be tested smoothly on ***test.xls***. One way to ensure this is to use the python package **argparse**. For example, we will require the code of group 25 to be able to run like this:

- `python3 group_25.py --test_set validate.xls`

Where `group_25.py` is group 25's main code file, and the variable `--test_set` determines which 2 years of price data to test your algorithm on. The above code snippet should thus run your algorithm on the validation set ***validate.xls***, and print a figure showing the accumulated profit over time. Next, your code should also be able to run:

- `python3 group_25.py --test_set test.xls`

In order to run the algorithm on the test set ***test.xls****, and* print a figure showing the accumulated profit over time. Note that this shouldn't be difficult, as ***validate.xls*** and ***test.xls*** are of the exact same dimensions/format.