Assignment 1 concentrates on renewable energy trading in day-ahead and balancing markets, following the lectures and tutorials from the first 5-6 weeks of the course. The work to undertake involves a bit of mathematical modelling, implementation in your favorite modelling language (preferably Python), generation and discussion of results, as well as presentation of the work in a short report.

The aim of Assignment 1 is to evaluate

- your understanding of day-ahead and balancing electricity markets,
- your ability to use real-world data as input,
- your ability to implement linear and non-linear regression with regularization,
- your ability to use regression indirectly and directly for trading,
- your ability to use unsupervised learning for better prediction,
- your critical analysis of the results generated.

The expected outcome of Assignment 1 includes:

- a report of maximum 12 pages (excluding appendices),
- code delivered as supplementary material.

The evaluation of Assignment 1 will count for 33% of the final grade. **Individual contributions to the assignment must be clearly stated in the report.** If not, equal contribution will be assumed. Good luck!

Description of the assignment

In this assignment, you will apply linear and non-linear regression with regularization for trading in day-ahead and balancing markets. Envision yourself as the owner of the wind farm, aiming to trade energy into electricity markets by learning from historical data. The historical data shared with you for this assignment is the actual wind power data for four wind farms in Bornholm, climate data of close locations, day-ahead and balancing prices. You need to select one of the wind farms along with its corresponding time-linked data. Note that Bornholm is in the DK2 market area, and you may disregard any instances of negative prices.

The goal of this assignment is to develop two models to be used for trading in day-ahead and balancing markets and investigate which one is better. In the initial approach, we employ regression to predict wind power, which we subsequently apply to address the decision-making problem related to offering strategies. In contrast, in the second approach, we utilize regression to directly determine the most effective offering strategy. Let us start with the first model and then proceeding to the second one.

Model 1:

Step 1) Trading optimization model: You have learnt how to develop an optimization model for renewable energy trading in day-ahead and balancing markets.

Step 1.1) Please write the corresponding optimization problem as a linear program.

Step 1.2) Please denote decision variables and parameters of this optimization problem.

Step 2) Data collection: In order to implement the optimization problem in Step 1, you need to predict the wind power over the next day and then use it as an input to the problem. From the course, you have learnt how to use regression as a means of making predictions. For that, you need to first construct the dataset using the data provided with this assignment. As potential suggestions, you may include features such as the actual wind power of the previous day, the 0.05, 0.5, 0.95 quantiles of wind production in the previous week, forecast date, climate conditions (e.g., wind speed, temperature, etc.). Please note that you are not restricted to the mentioned features and feel free to select the features you believe will enhance prediction accuracy. Additionally, not all the attached data needs to be used. It might be a good idea to start with the low number of features and samples.

Step 2.1) Please describe the number and types of features that you have chosen and explain how you can scale them.

Step 2.2) Now, you need to evaluate the performance of a predictor. To this end, please explain how you build up your training and testing dataset and what they include.

Step 3) Linear regression:

After constructing the training dataset, the next step is to train a regression model and use it for prediction. Let us start with the simplest one which is linear regression. As you know, there are two ways to solve the linear regression problem: implementing the gradient descent algorithm and using the closed form solution.

Step 3.1) Please show that these two methods end up with the same solution. For simplicity, you can start with the low number of samples, e.g. 100 datapoints.

Step 3.2) Please increase the number of samples to improve the accuracy of prediction and only use the closed form solution to find the optimal regression model.

Step 3.3) Verify your model using the testing dataset and appropriate evaluation metrics (e.g., Root mean squared error, Mean absolute error, R-squared).

Step 4) Nonlinear regression:

This step is an extension of the previous step. From the course, you have gained insights into transitioning from linear to non-linear regression.

Step 4.1) Please extend the linear regression model from Step 1 to a nonlinear model by modifying the features.

Step 4.2) Please extend the linear regression model from Step 1 to a nonlinear model by defining weights to the data points.

Step 4.3) Verify your models similar to step 3.3, and compare them with the linear regression model.

Step 5) Regularization:

Regularization is a fundamental concept in machine learning that helps prevent overfitting and improve generalization to unseen data. In the context of this assignment, we will focus on two common types of regularization: L1 (Lasso) and L2 (Ridge) regularization. These techniques will help us find a balance between fitting the training data perfectly and avoiding excessive complexity in our models.

Here's what you need to do:

Step 5.1) Modify your linear and non-linear regression models from Step 3 to include L1 regularization.

Step 5.2) Introduce a regularization parameter which controls the strength of regularization. Experiment with different values for the regularization parameter to find the optimal level of regularization.

Step 5.3) Train the regularized model on your training dataset and verify the model similar to step 3.3.

Step 5.4) Evaluate the performance of the L1-regularized model on your testing dataset using the same evaluation metrics as in Step 3.3 and Step 4.3.

Step 5.5) Analyze how regularization affects the coefficients of your features. Do some features become more influential, while others are effectively shrunk towards zero?

Step 5.6) Repeat Steps 5.1 to 5.5 with L2 regularization.

Step 6) Revenue calculation for evaluation:

The aim of this step is to evaluate the performance of your prediction by quantifying the revenue generated from day-ahead and balancing markets. Now, let us revisit the optimization problem outlined in Step 1.

Step 6.1) Given all regression models that you obtained from previous steps to predict the production of your wind farm, please calculate the expected revenue by using the historical day-ahead and balancing prices provided with this assignment.

Step 6.2) Which model works best to get higher revenue.

Step 7) Suggestion for improvement (This is an optional step with additional grade):

One potential way to improve the accuracy of your prediction is to use different regression models for different groups of data points. Here is what you can do for this purpose:

Step 7.1) Please use k-means clustering to cluster your datapoints into different groups.

Step 7.2) For each group of datapoints, you can try to apply different local regression models that you previously investigated. Evaluate whether this strategy enhances prediction accuracy and discuss your findings.

Model 2:

In this model, we leverage regression to **directly** determine the most effective offering strategy.

Step 1) Please adjust your training and testing dataset in terms of features and target, accordingly. To provide some guidance, note that the label in these datasets is not wind power production; instead, it represents the optimal offering strategy. To obtain this, you need to run the optimization problem from step 1 for each datapoint. Therefore, you may also need to add or remove relevant features.

Step 2) Please investigate whether regression or classification techniques work better in this context. Then, pick one suitable technique and build up Model 2.

Step 3) Compare the performance of Models 2 with 1, and summarize your findings.

Step 4) Can you provide any recommendations for further improving trading strategies based on your analysis?