

## Policies

- Prepare a single Jupyter notebook (`.ipynb`) containing your complete report, all executable code, and any generated outputs (e.g., images, results).
- Name your file “`Homework-2-YourName-YourStudentNumber.ipynb`” and submit it to Canvas.
- Submissions after the deadline will not be accepted.
- **Policy on AI Tools:**
  - The submission of any code, text, or answers directly generated by AI tools (including but not limited to ChatGPT, Claude, Copilot, or similar systems) as one’s own original work is **strictly prohibited**.
  - Using AI to complete the core, analytical, or interpretative parts of the assignment for you is **strictly prohibited**.
  - You are **allowed** to use AI as a search engine or coding assistant for tasks like explaining a general programming concept or error message; or looking up the syntax for a standard function (e.g., `sklearn.metrics.roc_curve`).
  - If you use an AI tool for any assistance, you must include a section titled “AI Use Declaration” at the end of your report. In this section, you must state which tool you used and describe specifically how you used it (e.g., “I used ChatGPT to debug a dimension mismatch error in my TensorFlow model” or “I used Claude to get an example of how to use `xgboost.plot_importance`”). **Failure to declare the use of AI will be treated as an academic integrity violation.**

## 1 Time series prediction [10 Points]

Time series prediction is a typical task in various scientific fields. This task focuses on developing a neural network-based model to predict the future evolution of a real-world energy consumption time series. The provided dataset comprises  $T = 242551$  hourly observations recorded between October 2004 and August 2018 (See dataset `EnergyConsumption_hourly.csv`).

**Problem A [6 points]:** Develop a neural network to predict energy consumption for January to August 2018. The model should be trained on all data preceding this period (i.e., the full series excluding the data points within January to August 2018). The performance goal is to minimize the Symmetric Mean Absolute Percentage Error (sMAPE):

$$\text{sMAPE} = \frac{2}{n} \sum_{t=T-n+1}^T \frac{|\hat{x}(t) - x(t)|}{|\hat{x}(t)| + |x(t)|} \times 100\%,$$

where  $n$  is the number of data points to be predicted,  $\hat{x}(t)$  and  $x(t)$  are the predicted and the real values at time point  $t$ , respectively.

Draw a simple picture to show the pipeline of your model, and tell the minimal sMAPE achieved by your model.

**Problem B [4 points]:** If the task is to predict energy consumption for January 2017 to August 2018, how to improve your model in Problem A for making such longer predictions? Tell the minimal sMAPE you achieved for this case.

## 2 Machine learning phases of matter [10 Points]

Two-dimensional Ising model is a classic model in physics, where we take a square lattice of points with a magnet at each point. These magnets can either spin up or down, i.e. each magnet can be in one of two states (+1 or -1). With interactions between magnets being short ranged, each magnet interacts only with its four neighbors. If the temperature is low enough the magnets gradually form bigger and bigger islands where they are aligned, and if we wait long enough they all become aligned; If the temperature is high enough the system never uniformizes because there is a balance between wanted flips and unwanted thermal flips. There is a **phase transition** between those two possible outcomes, happening at a known critical temperature. In this homework, given the pictures of the system states, you are asked to develop a neural network based method to tell whether the system is above or below the critical temperature.

See the reference for more information: J. Carrasquilla, R. Melko, Nature Physics (2017)  
<https://www.nature.com/articles/nphys4035>

You can use `Ising_simulation.ipynb` to generate training and test datasets which contain pictures at different temperatures.

**Problem A [4 points]:** Develop a neural network model, train your model using the training data, and then tell whether each of the pictures in the test dataset is above or below the critical temperature. In your

report, draw the pipeline of your model; plot the test accuracy vs. temperature (What do you observe?)

**Problem B [6 points]:** Develop a neural network model, train your model using the training data, and then predict the temperature for each of the pictures in the test dataset. In your report, draw the pipeline of your model; plot the test MAE (Mean Absolute Error) vs. temperature. What do you observe?