Deep Learning (FSS22)
Assignment 3: Hyperparameter Optimization

The archive provided to you contains this assignment description, datasets, as well as Python code fragments for you to complete. Comments and documentation in the code provide further information.

It suffices to fill out the "holes" that are marked in the code fragments provided to you, but feel free to modify the code to your liking. You need to stick with Python though.

Please adhere to the following guidelines in all the assignments. If you do not follow those guidelines, we may grade your solution as a FAIL.

Provide a single ZIP archive with name `dl22-<assignment number>-<your ILIAS login>.zip`. The archive needs to contain:

- A single PDF report that contains answers to the tasks specified in the assignment, including helpful figures and a high-level description of your approach. Do not simply convert your Jupyter notebook to a PDF! Write a separate document, stay focused and brief. As a guideline, try to stay below 10 pages.
- All the code that you created and used in its original format.
- A PDF document that renders your Jupyter notebook with all figures. (If you don't use Jupyter, then you obviously do not need to provide this.)

A high quality report is required to achieve an EXCELLENT grade. Such a report is self-explanatory (i.e. do not refer to your code except for implementation-only tasks), follows standard scientific practice (e.g. when using images, tables or citations), does not include on hand-written notes, and does not exceed 10 pages. In addition, label all figures (and refer to figure labels in your write-up), include references if you used additional sources or material, and use the tasks numbers of the assignments as your section and subsection numbers.

Hand-in your solution via ILIAS until the date specified there. This is a hard deadline.

1 Basic Hyperparameter Tuning

In this task, we aim to minimize a function $f_1 : [0, 1]^2 \to \mathbb{R}$ using basic hyperparameter tuning. **Do not look at this function** before you solved and wrote down your solution of subtask a).

Do not use any HPO library for this task, but implement everything yourself.

a) Perform a manual hyperparameter search with 12 trials. The provided code already sets the first trial and provides facilities to visualize your trials so far. Try to obtain an as good function value as possible. Which results did you get? What was your reasoning? Do you think you found a good solution?

   Note. We do not subtract points if you do not obtain a decent result. So, no need to cheat.

b) Plot function f_1 using the provided code. Compare to your results of a). Discuss.

c) Perform a grid search with 12 trials. Discuss.

d) Perform a random search with 12 trials (i.i.d., uniform). Repeat multiple times and discuss.

   Optional. Repeat this task using a Sobol sequence instead of random trials (see provided code).

e) Suppose we are optimizing a function $f : [0, 1]^D \to \mathbb{R}$ in the D-dimensional unit cube. We perform n trials using random search (i.i.d., uniform). How many trials do we need such that with high probability, we find at least one good HP configuration when

   (i) a fraction of $0 \leq p \leq 1$ of the HP configuration space is good,

   (ii) a fraction of $0 \leq p \leq 1$ of each dimension of the HP configuration space is good (and all other choices along the dimension lead to bad configurations)?

   Discuss. What if we used grid search instead of random search?

2 Gaussian Process Regression

In this task, we experiment with Gaussian Process Regression (GPR) to estimate function $f_2 : \mathbb{R} \to \mathbb{R}$. The function, as well as auxiliary information and potentially helpful pieces of code are given provided Python files.

a) We are using scaled RBF kernels in this task. Experiment with the scale and length scale parameters of this kernel. How do they affect the kernel function? Intuitively, what do these parameters control?

b) Experiment with GP priors for different parameterizations of the kernel. What is the influence of the kernel and its parameters?

   Optional. Scikit-learn's Gaussian Process model does not allow specification of a mean function. Which mean function is it using? Can we nevertheless use the model if we wanted to use an alternative mean function?

c) Plot f_2. Use your insight from the previous subtasks to select a suitable scale and length scale parameter for the kernel of a GP prior. Argue for your choice. (Note that this is generally not possible in practice since the function to be optimized is unknown. We do it here to gain intuition.)

   Perform Gaussian Process Regression using $n = 3$ data points of f_2 using your kernel parameters. Plot the posterior distribution and discuss. Was your choice suitable?

d) Experiment with different kernel parameters and number of training examples. Discuss your findings.

e) Suppose we want to use GPR for Bayesian Optimization. Discuss your findings in this context.

3 Bayesian Optimization

Explore the available libraries for HPO. Pick a library that implements Bayesian Optimization. Use the library (with default settings) to optimize f_1 and f_2. Study the explored configurations and discuss.

Optional. Experiment with other HPO methods of your choice.