BlackBox CapStone project

Richard Meyer, August 2023

#### A blue circle with green outline and three dots in the middle Description automatically generated **Directions**

You now have all the tools you need to get started on your capstone project. In the [previous video](https://classroom.emeritus.org/courses/2993/pages/capstone-strategies), Ruth presented some options for identifying and refining code that can be used for tuning hyperparameters in machine learning models. In this reflection activity, think about how you’ll approach the competition for this programme. Your submission should answer the following:

* Of the three personas covered in the video, which best describes the strategy you will be using to win the competition?
* From where will you obtain your initial codebase?
* How does the strategy you’ve identified align with your career goals and objectives following completion of the programme?

# Overall approach

Very much “trying to come up with an improved algorithm”, and will definitely code it myself as much as possible.

I believe this is the actual problem to be solved: <http://bbochallenge.com>

# Key Steps

## Preliminary steps

* Understand what the actual challenge is, and any constraints (number of iterations for example).
* Find out whether any collaboration with other course participants is possible.
* Research existing solutions: investigate the corresponding papers & videos.
* Read any existing code to see how much I understand.
* Learn more online using google, WikiPedia etc.

## Exploration

* Code a basic approach myself and see how it performs.
* **A key data point will be how expensive the BlackBox is to evaluate.**
* Try some naïve approaches such as random parameters or grid search for example.
* Try to understand where the difficulty comes from:
  + Heteroscedasticity poses a challenge.
  + Scaling parameter.
* Extract plots of some of the variables and their impact on the overall metrics.

## Iterative Probem Solving

* Try different algorithms to address the challenges above.
* Examine the data produced.
* Generate new ideas and iterate…
* Consider a Gaussian Process to optimise the Gaussian Process used to tune the hyper-parameters. ☺

## PUBLISHED Solutions

### Huawei

Overview of their solution with links to their code-base and a video presentation.

<https://www.reddit.com/r/MachineLearning/comments/ll94ag/r_how_we_won_the_neurips_2020_black_box/>

Paper: <https://valohaichirpprod.blob.core.windows.net/papers/huawei.pdf>

### Nvidia

Have published paper explaining their approach:

<https://valohaichirpprod.blob.core.windows.net/papers/nvidia.pdf>

The code is available here: <https://github.com/daxiongshu/rapids-ai-BBO-2nd-place-solution>

**However, I suspect that this solution will be effectively unaffordable in terms of the hardware required, though I can get some GPU time on google colab.**

They quote: “On a DGX-1, the search time is **reduced from more than 10 days on two 20-core CPUs to less than 24 hours on 8-GPUs.”**

**Original: 10 x 24 x 2 x 20 = 19,200 CPU-core-hours**

**GPU: 24 x 8 = 192 GPU hours**

**Ratio = 100 – sounds plausible, ie: I would have expected a minimum ratio of 30.**

### **JetBrains**

**git repository here** <https://github.com/jbr-ai-labs/bbo-challenge-jetbrains-research>**, with a paper available here:** <https://arxiv.org/pdf/2012.10335.pdf>

## Possible Algorithms

The best approach may be to code a Gaussian Process first, then see what can be done to further refine the process.

# Personal Goals

Clearly taking part in this challenge will be an interesting exercise, personal objectives include:

* Gaining a better knowledge of Hyper Parameter optimisation and its inherent challenges.
* Improving my Python and ML skills.
* Reading a number of papers on the topic and increasing the amount I understand (currently 30% at best, but that’s up from 10% a couple of months ago).
* Possibly creating a re-usable codebase for future projects.

We can hopefully have an interesting debrief with the other students to see how they handled the problem and how they won!