

STAT40830 - Adv Data Prog with R (online)

Assignment 2

Due on Wednesday 22nd July 2020 11:59pm (IST).

This assignment is worth 20% of your overall grade for this module. You should carry out your analysis in R and submit a compressed folder of all source files to Brightspace before the assignment deadline.

Instructions

For this assignment you should modify the `climr` package we created in Week 6 to include:

- a new method for the class `"climr"` called `gp_fit`. This function takes in input the class object and a string argument, which should be one of `c("Nelder-Mead", "BFGS", "SANN", "Brent")`. The method `gp_fit` should be an adaptation of the function `regression_fit` which we have used in Assignment 1, and should optimise the hyperparameters of the Gaussian process using `optim` with the optimisation method given by the string argument above. The method should return an object of class `climr_gp_fit`.
- a plot method for the class `climr_gp_fit`. Similarly to the plot method for a `climr_fit` object, this method will use `ggplot2` to illustrate the data as a scatter-plot, and will add the smoothed Gaussian process regression line. See Figure 1 at the end of this document for reference.

Once the new package is loaded, the following code should be used to produce the plot attached:

```
data1 <- load_clim("NH")
data2 <- gp_fit(data1, "BFGS")
plot(data2)
```

Note:

- complete the assignment using all the indications given throughout the first 6 weeks of the module.
- the `regression_fit` function that we have used in Assignment 1 works on standardised data. So, in your code, you should do and undo the data rescaling as needed.
- the style of coding should be sensible and easy to read and mark.
- the package should contain all the appropriate documentation.
- the package should be built and checked with 0 errors, 0 warnings and 0 notes (some notes and warnings may be system-dependent so certain exceptions may be accepted on this point).
- you should submit only the package sources (i.e. the .tar.gz file).
- there is no grading scheme as the questions are rather open and the assignment may be solved in a number of ways. Full marks will be awarded if there are no errors/warnings/notes, the example above produces the plot as requested, and the package is well documented/commented.

Please ensure that all documents are uploaded to Brightspace **clearly and on time**. It is students' responsibility in taking care of uploading all assignments. Pending or incomplete submissions after the deadline will be considered late and be penalised according to UCD guidelines, unless proven and valid justification (e.g. doctor's certificate) is provided. **Plagiarism is prohibited**. Please refer to the UCD Plagiarism Policy.

In addition, you must not discuss your answers or attempts on any of the module forums or with your classmates. Questions relating to the assignment will not be answered as the work is assessed.

Good luck and enjoy the assignment.

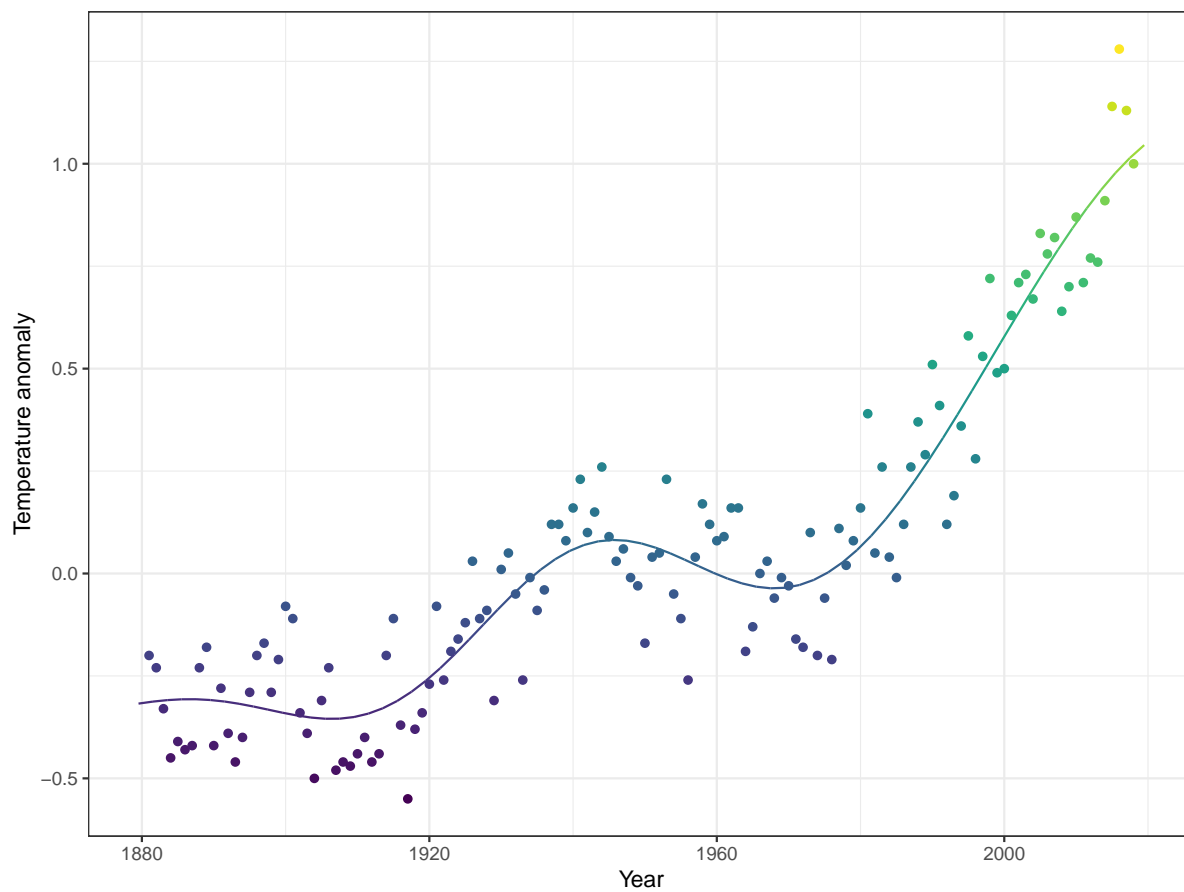


Figure 1: Example `ggplot` that your package should produce.