Backend ML Engineer Case Study

August 15, 2022

In this case study, you will be implementing an algorithm called “stacking” for survival analysis. The goal of this algorithm is to allow us to use conventional classification models in the context of survival analysis. The intention here is for you to implement the code in a way that mimics the way that you would implement it on the job as closely as possible. The instructions are as follows:

1. Read sections 1 and 2 of the attached paper titled *Stacking paper.pdf*.
2. In a file called *stacking.py*, design an API for a Python class that will allow the client code to both fit a survival model using the algorithm described in the paper as well as apply the model to unseen data. Some questions you may want to consider while doing this are:
   1. What methods might be useful for client code to be able to call?
   2. Are there any other libraries that you can look to for inspiration for API design? Are there any conventions that you may want to respect?
   3. How can you provide flexibility for client code to select which underlying classification model to use?
3. Create a toy data set on which you may test your implementation.
4. In a file called *test\_stacking.py*, write unit tests for your class. This should include at least one test that uses your toy data set.
5. Create a *README.md* file containing documentation for how to invoke your new class.
6. Implement the API that you have defined in step 2 in Python.
7. Zip *README.md*, *stacking.py*, *test\_stacking.py,* and the file containing your toy data set and send them back to us.

Note that if you should feel free to send us any clarifying questions that you may have as you implement the desired functionality.