A Note on Pickling

More about 'pickling'

There are several popular ways to save (and finalize) a model. To name some, you can use Joblib (a part of the SciPy ecosystem), and JSON. Certainly, each of those choices has its pros and cons. Pickle is probably the most intuitive and definitely our preferred choice.

Once again, 'pickle' is the standard Python tool for serialization and deserialization. In simple words, pickling means: converting a Python object (no matter what) into a string of characters. Logically, unpickling is about converting a string of characters (that has been pickled) into a Python object.

There are some potential issues you should be aware of, though!

1. ***Pickle and Python version.*** Pickling is strictly related to Python version. It is **not recommended** to (de)serialize objects across different Python versions. Logically, if you're working on your own this will never be an issue (unless you upgrade/downgrade your Python version).

2. ***Pickle is slow.*** Well, you will barely notice that but for complex structures it may take loads of time to pickle and unpickle.

3. ***Pickle is not secure.*** This is evident from the documentation of pickle, quote: "Never unpickle data received from an untrusted or unauthenticated source." The reason is that just about anything can be pickled, so you can easily unpickle malicious code.

   Now, if you are unpickling your own code, you are more or less safe.

   If, however, you receive pickled objects from someone you don't fully trust, you should be very cautious. That's how viruses affect your operating system.

   Finally, even your own file may be changed by an attacker. Thus, the next time you unpickle, you can unpickle just about anything (that this unethical person put there).

Certainly, all these cases are very rare, but you must be aware of them. Generally, we recommend using JSON, but that's a topic for another time 😊

Stay safe!