I actually recently had the same question. You can see a big difference submitting to the diagnostics as well. The downloaded file has .25 corr with .95 sharpe but the file produced from the example_model.py

is much worse than that.