

March 19, 2023

0.1 *Out-of-core machine learning*

Out-of-core machine learning is a technique used to **handle large datasets** that cannot fit into memory.

0.1.1 *1.Stream Data:*

In out-of-core machine learning, the data is read from disk in small chunks, also known as batches, and processed sequentially. This process is often called streaming, as the data is processed in a continuous stream. The size of the batches is typically small enough to fit into memory, but large enough to be efficient.

0.1.2 *2.Extract Features:*

Once the data has been read into memory, the next step is to extract features from the data. Feature extraction is the process of transforming the raw data into a set of meaningful features that can be used to train the model. This step is critical, as the quality of the features will directly impact the accuracy of the model.

0.1.3 *3.Train Model:*

Once the features have been extracted, the model is trained on the data. The training process typically involves adjusting the model's parameters to minimize the difference between the predicted output and the actual output.

```
[ ]: import pandas as pd
```

```
[ ]: df = pd.read_csv('large_dataset.csv')
```

```
[ ]: df.shape
```

```
[ ]: (30000000, 5)
```

```
[ ]: df.head()
```

```
[ ]:      f1      f2      f3      f4      target
0  0.203936 -0.291191 -0.617251 -1.461300  10.796660
1 -1.195432  0.069214  0.291861  1.972511  11.154018
2 -2.560271 -0.827846  0.405551 -0.132588 -20.342684
3 -1.131196  0.252654  0.961175  1.236873   1.407638
```

```
4 0.858774 -0.461605 1.095498 -0.871580 -17.043060
```

```
[ ]: df.describe()
```

```
[ ]:
      f1      f2      f3      f4      target
count  3.000000e+07  3.000000e+07  3.000000e+07  3.000000e+07  3.000000e+07
mean   -3.258116e-04 -1.060974e-04 -1.360234e-05 -6.534420e-05 -8.363976e-03
std     9.999900e-01  1.000074e+00  9.998615e-01  1.000105e+00  3.131135e+01
min    -5.491886e+00 -5.333572e+00 -5.312364e+00 -5.237933e+00 -1.694348e+02
25%    -6.748877e-01 -6.744943e-01 -6.743856e-01 -6.744510e-01 -2.112416e+01
50%    -3.116059e-04 -2.164725e-04 -4.358653e-05 -2.934173e-04 -1.129371e-02
75%     6.741835e-01  6.743107e-01  6.744813e-01  6.746014e-01  2.111568e+01
max     5.500586e+00  5.449671e+00  5.500811e+00  5.347441e+00  1.778399e+02
```

```
[ ]: !pip install vaex
```

```
Requirement already satisfied: vaex in c:\users\91842\anaconda3\lib\site-
packages (4.4.0)
Requirement already satisfied: vaex-jupyter<0.7,>=0.6.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.6.0)
Requirement already satisfied: vaex-hdf5<0.10,>=0.9.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.9.0)
Requirement already satisfied: vaex-viz<0.6,>=0.5.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.5.0)
Requirement already satisfied: vaex-server<0.7,>=0.6.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.6.0)
Requirement already satisfied: vaex-astro<0.9,>=0.8.3 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.8.3)
Requirement already satisfied: vaex-core<5,>=4.4.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (4.4.0)
Requirement already satisfied: vaex-ml<0.14,>=0.13.0 in
c:\users\91842\anaconda3\lib\site-packages (from vaex) (0.13.0)
Requirement already satisfied: astropy in c:\users\91842\anaconda3\lib\site-
packages (from vaex-astro<0.9,>=0.8.3->vaex) (4.2.1)
Requirement already satisfied: progressbar2 in
c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex)
(3.53.1)
Requirement already satisfied: future>=0.15.2 in
c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex)
(0.18.2)
Requirement already satisfied: six in c:\users\91842\anaconda3\lib\site-packages
(from vaex-core<5,>=4.4.0->vaex) (1.15.0)
Requirement already satisfied: cloudpickle in c:\users\91842\anaconda3\lib\site-
packages (from vaex-core<5,>=4.4.0->vaex) (1.6.0)
Requirement already satisfied: tabulate>=0.8.3 in
c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex)
(0.8.9)
```

Requirement already satisfied: frozendict in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (2.0.6)

Requirement already satisfied: dask in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (2021.4.0)

Requirement already satisfied: pyyaml in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (5.4.1)

Requirement already satisfied: pandas in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (1.2.4)

Requirement already satisfied: aplus in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (0.11.0)

Requirement already satisfied: blake3 in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (0.2.0)

Requirement already satisfied: numpy>=1.16 in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (1.20.1)

Requirement already satisfied: nest-asyncio>=1.3.3 in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (1.5.1)

Requirement already satisfied: requests in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (2.25.1)

Requirement already satisfied: pyarrow>=3.0 in c:\users\91842\anaconda3\lib\site-packages (from vaex-core<5,>=4.4.0->vaex) (5.0.0)

Requirement already satisfied: h5py>=2.9 in c:\users\91842\anaconda3\lib\site-packages (from vaex-hdf5<0.10,>=0.9.0->vaex) (2.10.0)

Requirement already satisfied: ipyvuetify<2,>=1.2.2 in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (1.8.1)

Requirement already satisfied: ipyvolume>=0.4 in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (0.5.2)

Requirement already satisfied: ipyleaflet in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (0.14.0)

Requirement already satisfied: bqplot>=0.10.1 in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (0.12.30)

Requirement already satisfied: ipympl in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (0.7.0)

Requirement already satisfied: xarray in c:\users\91842\anaconda3\lib\site-packages (from vaex-jupyter<0.7,>=0.6.0->vaex) (0.19.0)

Requirement already satisfied: traittypes>=0.0.6 in c:\users\91842\anaconda3\lib\site-packages (from bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.2.1)

Requirement already satisfied: traitlets>=4.3.0 in c:\users\91842\anaconda3\lib\site-packages (from bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (5.0.5)

Requirement already satisfied: ipywidgets>=7.5.0 in c:\users\91842\anaconda3\lib\site-packages (from bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (7.6.3)

Requirement already satisfied: ipywebtrc in c:\users\91842\anaconda3\lib\site-packages (from ipyvvolume>=0.4->vaex-jupyter<0.7,>=0.6.0->vaex) (0.6.0)

Requirement already satisfied: pythreajs>=1.0.0 in c:\users\91842\anaconda3\lib\site-packages (from ipyvvolume>=0.4->vaex-jupyter<0.7,>=0.6.0->vaex) (2.3.0)

Requirement already satisfied: Pillow in c:\users\91842\anaconda3\lib\site-packages (from ipyvvolume>=0.4->vaex-jupyter<0.7,>=0.6.0->vaex) (8.2.0)

Requirement already satisfied: ipyvue<2,>=1.5 in c:\users\91842\anaconda3\lib\site-packages (from ipyvuetify<2,>=1.2.2->vaex-jupyter<0.7,>=0.6.0->vaex) (1.5.0)

Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\users\91842\anaconda3\lib\site-packages (from ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (3.5.1)

Requirement already satisfied: ipython>=4.0.0 in c:\users\91842\anaconda3\lib\site-packages (from ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (7.22.0)

Requirement already satisfied: nbformat>=4.2.0 in c:\users\91842\anaconda3\lib\site-packages (from ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (5.1.3)

Requirement already satisfied: ipykernel>=4.5.1 in c:\users\91842\anaconda3\lib\site-packages (from ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (5.3.4)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\91842\anaconda3\lib\site-packages (from ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (1.0.0)

Requirement already satisfied: tornado>=4.2 in c:\users\91842\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (6.1)

Requirement already satisfied: jupyter-client in c:\users\91842\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (6.1.12)

Requirement already satisfied: prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0 in c:\users\91842\anaconda3\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (3.0.17)

Requirement already satisfied: decorator in c:\users\91842\anaconda3\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (5.0.6)

Requirement already satisfied: backcall in c:\users\91842\anaconda3\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.2.0)

Requirement already satisfied: colorama in c:\users\91842\anaconda3\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.4.4)

Requirement already satisfied: pygments in c:\users\91842\anaconda3\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-

```

jupyter<0.7,>=0.6.0->vaex) (2.8.1)
Requirement already satisfied: jedi>=0.16 in c:\users\91842\anaconda3\lib\site-
packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.17.2)
Requirement already satisfied: setuptools>=18.5 in
c:\users\91842\anaconda3\lib\site-packages (from
ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (52.0.0.post20210125)
Requirement already satisfied: pickleshare in c:\users\91842\anaconda3\lib\site-
packages (from ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.7.5)
Requirement already satisfied: parso<0.8.0,>=0.7.0 in
c:\users\91842\anaconda3\lib\site-packages (from
jedi>=0.16->ipython>=4.0.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.7.0)
Requirement already satisfied: jupyter-core in
c:\users\91842\anaconda3\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (4.7.1)
Requirement already satisfied: ipython-genutils in
c:\users\91842\anaconda3\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.2.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
c:\users\91842\anaconda3\lib\site-packages (from
nbformat>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (3.2.0)
Requirement already satisfied: attrs>=17.4.0 in
c:\users\91842\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbform
at>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex)
(20.3.0)
Requirement already satisfied: pyrsistent>=0.14.0 in
c:\users\91842\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbform
at>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex)
(0.17.3)
Requirement already satisfied: wcwidth in c:\users\91842\anaconda3\lib\site-
packages (from prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipy
widgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.2.5)
Requirement already satisfied: ipydatawidgets>=1.1.1 in
c:\users\91842\anaconda3\lib\site-packages (from
pythreejs>=1.0.0->ipyvolume>=0.4->vaex-jupyter<0.7,>=0.6.0->vaex) (4.2.0)
Requirement already satisfied: jinja2 in c:\users\91842\anaconda3\lib\site-
packages (from vaex-ml<0.14,>=0.13.0->vaex) (2.11.3)
Requirement already satisfied: numba in c:\users\91842\anaconda3\lib\site-
packages (from vaex-ml<0.14,>=0.13.0->vaex) (0.53.1)
Requirement already satisfied: fastapi[all] in
c:\users\91842\anaconda3\lib\site-packages (from vaex-server<0.7,>=0.6.0->vaex)
(0.68.0)

```

Requirement already satisfied: cachetools in c:\users\91842\anaconda3\lib\site-packages (from vaex-server<0.7,>=0.6.0->vaex) (4.2.2)

Requirement already satisfied: matplotlib>=1.3.1 in c:\users\91842\anaconda3\lib\site-packages (from vaex-viz<0.6,>=0.5.0->vaex) (3.3.4)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\91842\anaconda3\lib\site-packages (from matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.0->vaex) (1.3.1)

Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in c:\users\91842\anaconda3\lib\site-packages (from matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.0->vaex) (2.4.7)

Requirement already satisfied: python-dateutil>=2.1 in c:\users\91842\anaconda3\lib\site-packages (from matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.0->vaex) (2.8.1)

Requirement already satisfied: cycler>=0.10 in c:\users\91842\anaconda3\lib\site-packages (from matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.0->vaex) (0.10.0)

Requirement already satisfied: notebook>=4.4.1 in c:\users\91842\anaconda3\lib\site-packages (from widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (6.3.0)

Requirement already satisfied: Send2Trash>=1.5.0 in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (1.5.0)

Requirement already satisfied: nbconvert in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (6.0.7)

Requirement already satisfied: pyzmq>=17 in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (20.0.0)

Requirement already satisfied: prometheus-client in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.10.1)

Requirement already satisfied: terminado>=0.8.3 in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.9.4)

Requirement already satisfied: argon2-cffi in c:\users\91842\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (20.1.0)

Requirement already satisfied: pywin32>=1.0 in c:\users\91842\anaconda3\lib\site-packages (from jupyter-core->nbformat>=4.2.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (227)

Requirement already satisfied: pywinpty>=0.5 in c:\users\91842\anaconda3\lib\site-packages (from terminado>=0.8.3->notebook>=4.4

```
.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.5.7)
Requirement already satisfied: cffi>=1.0.0 in c:\users\91842\anaconda3\lib\site-
packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidg-
ets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (1.14.5)
Requirement already satisfied: pycparser in c:\users\91842\anaconda3\lib\site-
packages (from cffi>=1.0.0->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.
5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (2.20)
Requirement already satisfied: pyerfa in c:\users\91842\anaconda3\lib\site-
packages (from astropy->vaex-astro<0.9,>=0.8.3->vaex) (1.7.3)
Requirement already satisfied: fsspec>=0.6.0 in
c:\users\91842\anaconda3\lib\site-packages (from dask->vaex-
core<5,>=4.4.0->vaex) (0.9.0)
Requirement already satisfied: partd>=0.3.10 in
c:\users\91842\anaconda3\lib\site-packages (from dask->vaex-
core<5,>=4.4.0->vaex) (1.2.0)
Requirement already satisfied: toolz>=0.8.2 in
c:\users\91842\anaconda3\lib\site-packages (from dask->vaex-
core<5,>=4.4.0->vaex) (0.11.1)
Requirement already satisfied: locket in c:\users\91842\anaconda3\lib\site-
packages\loket-0.2.1-py3.8.egg (from partd>=0.3.10->dask->vaex-
core<5,>=4.4.0->vaex) (0.2.1)
Requirement already satisfied: starlette==0.14.2 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (0.14.2)
Requirement already satisfied:
pydantic!=1.7,!1.7.1,!1.7.2,!1.7.3,!1.8,!1.8.1,<2.0.0,>=1.6.2 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.8.2)
Requirement already satisfied: aiofiles<0.6.0,>=0.5.0 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (0.5.0)
Requirement already satisfied: email_validator<2.0.0,>=1.1.1 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.1.3)
Requirement already satisfied: itsdangerous<2.0.0,>=1.1.0 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.1.0)
Requirement already satisfied: orjson<4.0.0,>=3.2.1 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (3.6.2)
Requirement already satisfied: async_exit_stack<2.0.0,>=1.0.1 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.0.1)
Requirement already satisfied: async_generator<2.0.0,>=1.10 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.10)
Requirement already satisfied: ujson<5.0.0,>=4.0.1 in
```

```

c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (4.0.2)
Requirement already satisfied: graphene<3.0.0,>=2.1.8 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (2.1.9)
Requirement already satisfied: uvicorn[standard]<0.14.0,>=0.12.0 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (0.13.4)
Requirement already satisfied: python-multipart<0.0.6,>=0.0.5 in
c:\users\91842\anaconda3\lib\site-packages (from fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (0.0.5)
Requirement already satisfied: dnspython>=1.15.0 in
c:\users\91842\anaconda3\lib\site-packages (from
email_validator<2.0.0,>=1.1.1->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex)
(2.1.0)
Requirement already satisfied: idna>=2.0.0 in c:\users\91842\anaconda3\lib\site-
packages (from email_validator<2.0.0,>=1.1.1->fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (2.10)
Requirement already satisfied: aniso8601<=7,>=3 in
c:\users\91842\anaconda3\lib\site-packages (from
graphene<3.0.0,>=2.1.8->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (7.0.0)
Requirement already satisfied: graphql-relay<3,>=2 in
c:\users\91842\anaconda3\lib\site-packages (from
graphene<3.0.0,>=2.1.8->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (2.0.1)
Requirement already satisfied: graphql-core<3,>=2.1 in
c:\users\91842\anaconda3\lib\site-packages (from
graphene<3.0.0,>=2.1.8->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (2.3.2)
Requirement already satisfied: promise<3,>=2.3 in
c:\users\91842\anaconda3\lib\site-packages (from graphql-
core<3,>=2.1->graphene<3.0.0,>=2.1.8->fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (2.3)
Requirement already satisfied: rx<2,>=1.6 in c:\users\91842\anaconda3\lib\site-
packages (from graphql-core<3,>=2.1->graphene<3.0.0,>=2.1.8->fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (1.6.1)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\91842\anaconda3\lib\site-packages (from jinja2->vaex-
ml<0.14,>=0.13.0->vaex) (1.1.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
c:\users\91842\anaconda3\lib\site-packages (from pydantic!=1.7,!1.7.1,!1.7.2,!
=1.7.3,!1.8,!1.8.1,<2.0.0,>=1.6.2->fastapi[all]->vaex-
server<0.7,>=0.6.0->vaex) (3.7.4.3)
Requirement already satisfied: chardet<5,>=3.0.2 in
c:\users\91842\anaconda3\lib\site-packages (from requests->vaex-
core<5,>=4.4.0->vaex) (4.0.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
c:\users\91842\anaconda3\lib\site-packages (from requests->vaex-
core<5,>=4.4.0->vaex) (1.26.4)
Requirement already satisfied: certifi>=2017.4.17 in

```


c:\users\91842\anaconda3\lib\site-packages (from requests->vaex-core<5,>=4.4.0->vaex) (2020.12.5)

Requirement already satisfied: click==7.* in c:\users\91842\anaconda3\lib\site-packages (from uvicorn[standard]<0.14.0,>=0.12.0->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (7.1.2)

Requirement already satisfied: h11>=0.8 in c:\users\91842\anaconda3\lib\site-packages (from uvicorn[standard]<0.14.0,>=0.12.0->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (0.12.0)

Requirement already satisfied: watchdog>=0.6 in c:\users\91842\anaconda3\lib\site-packages (from uvicorn[standard]<0.14.0,>=0.12.0->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (0.7)

Requirement already satisfied: python-dotenv>=0.13 in c:\users\91842\anaconda3\lib\site-packages (from uvicorn[standard]<0.14.0,>=0.12.0->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (0.19.0)

Requirement already satisfied: websockets==8.* in c:\users\91842\anaconda3\lib\site-packages (from uvicorn[standard]<0.14.0,>=0.12.0->fastapi[all]->vaex-server<0.7,>=0.6.0->vaex) (8.1)

Requirement already satisfied: defusedxml in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.7.1)

Requirement already satisfied: jupyterlab-pygments in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.1.2)

Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.5.3)

Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.8.4)

Requirement already satisfied: bleach in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (3.3.0)

Requirement already satisfied: entrypoints>=0.2.2 in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.3)

Requirement already satisfied: testpath in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (0.4.4)

Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\91842\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-

```
jupyter<0.7,>=0.6.0->vaex) (1.4.3)
Requirement already satisfied: packaging in c:\users\91842\anaconda3\lib\site-
packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ip
ywidgets>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.7,>=0.6.0->vaex) (20.9)
Requirement already satisfied: webencodings in
c:\users\91842\anaconda3\lib\site-packages (from bleach->nbconvert->notebook>=4.
4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.7,>=0.6.0->vaex) (0.5.1)
Requirement already satisfied: llvmlite<0.37,>=0.36.0rc1 in
c:\users\91842\anaconda3\lib\site-packages (from numba->vaex-
ml<0.14,>=0.13.0->vaex) (0.36.0)
Requirement already satisfied: pytz>=2017.3 in
c:\users\91842\anaconda3\lib\site-packages (from pandas->vaex-
core<5,>=4.4.0->vaex) (2021.1)
Requirement already satisfied: python-utils>=2.3.0 in
c:\users\91842\anaconda3\lib\site-packages (from progressbar2->vaex-
core<5,>=4.4.0->vaex) (2.5.6)
```

```
[ ]: import vaex
```

```
[ ]: vaex_df = vaex.from_pandas(df)
```

```
[ ]: type(vaex_df)
```

```
[ ]: vaex.dataframe.DataFrameLocal
```

```
[ ]: vaex_df.export_hdf5('large_data.hdf5')
```

```
[ ]: vaex_df = vaex.open('large_data.hdf5')
```

```
[ ]: vaex_df.head()
```

```
[ ]:
#          f1          f2          f3          f4          target
0    0.203936 -0.291191 -0.617251 -1.4613      10.7967
1   -1.19543  0.0692141  0.291861  1.97251      11.154
2   -2.56027 -0.827846  0.405551 -0.132588     -20.3427
3   -1.1312  0.252654  0.961175  1.23687       1.40764
4    0.858774 -0.461605  1.0955   -0.87158     -17.0431
5    0.63273  2.07437   -0.700978  1.36085      49.2754
6   -1.92771  0.155743 -0.394158 -1.3712      11.0902
7   -0.165462 -0.481113 -1.29813  1.75204     -41.1136
8   -1.31883  0.393728 -1.83311 -0.599728      1.87724
9   -0.671377 -0.172073 -0.75585  1.05745     -26.2623
```

```
[ ]: vaex_df = vaex_df.shuffle()
df_train, df_test = vaex_df.ml.train_test_split(test_size=0.2)
```

```
C:\Users\91842\anaconda3\lib\site-packages\vaex\ml\__init__.py:31: UserWarning:
```

```
Make sure the DataFrame is shuffled
warnings.warn('Make sure the DataFrame is shuffled')
```

```
[ ]: df_train.shape
```

```
[ ]: (24000000, 5)
```

```
[ ]: df_test.shape
```

```
[ ]: (6000000, 5)
```


0.1.4 *Incremental Predictor Algorithms:*

- Incremental predictor algorithms ,also known as online learning algorithms that can update their predictions in real-time as new data becomes available.
- These algorithms are well-suited to out-of-core machine learning, as they can handle streaming data and can update the model in real-time.
- Some popular incremental predictor algorithms used in out-of-core machine learning include stochastic gradient descent (SGD), adaptive boosting (AdaBoost), and online random forests.

```
[ ]: from vaex.ml.sklearn import IncrementalPredictor
    from sklearn.linear_model import SGDRegressor

    features = ['f1','f2','f3','f4']
    target = 'target'

    model = SGDRegressor()

    vaex_model = 
    ↪ IncrementalPredictor(features=features,target=target,model=model,batch_size=500000)

    vaex_model.fit(df=df_train,progress='widget')
```

```
HBox(children=(FloatProgress(value=0.0, max=1.0), Label(value='In progress...')))
```

0.1.5 *Few Incremental Prediction Algorithms are:*

0.1.6 *1.Stochastic Gradient Descent (SGD):*

SGD is an iterative algorithm used for optimization in machine learning. It updates the model's parameters based on the error calculated from a single data point. SGD is widely used in out-of-core machine learning because it can handle large datasets and update the model in real-time.

0.1.7 *2.Adaptive Boosting (AdaBoost):*

AdaBoost is a boosting algorithm that combines multiple weak classifiers to form a strong classifier. It works by giving more weight to misclassified data points and less weight to correctly classified data points. AdaBoost is often used in out-of-core machine learning because it can update the model in real-time and handle streaming data.

0.1.8 3. Online Random Forests:

Random Forests are an ensemble learning technique that combines multiple decision trees to form a more accurate model. Online Random Forests are a variant of Random Forests that can update the model in real-time as new data becomes available. They are well-suited for out-of-core machine learning because they can handle large datasets and streaming data.

0.1.9 4. Passive Aggressive (PA) Algorithm:

PA is a family of online learning algorithms that are designed to handle binary classification problems. It works by making updates to the model's parameters based on the difference between the predicted output and the actual output. PA is often used in out-of-core machine learning because it can update the model in real-time and handle streaming data.

0.1.10 5. Vowpal Wabbit (VW):

VW is an open-source machine learning library that is designed to handle large datasets and out-of-core machine learning. It uses online learning algorithms, such as stochastic gradient descent and adaptive regularization, to train models on streaming data.

```
[ ]: df_test = vaex_model.transform(df_test)
```

```
[ ]: df_test.head()
```

```
[ ]: #          f1          f2          f3          f4      target      prediction
      0  0.651098    1.75828   -1.55602    0.205312    52.0572        41.9285
      1 -0.930079    0.66736    1.38524   -0.785785    17.9182        16.1285
      2 -0.757777   -1.0052    1.03499    0.227977   -37.9689       -24.1394
      3  0.320047    1.03793    1.27838    2.78425    -7.66551       25.0803
      4 -1.57124   -0.383367    0.162823   -0.70573   -13.792       -9.57404
      5 -1.25979   -0.777414    0.224138   -0.925356    21.0758      -18.9242
      6 -0.0639199 -0.143284   -0.103907   -0.761464   -9.42623       -3.48351
      7  0.798208    0.53451   -1.62919   -1.36567    -8.41418       12.6728
      8  0.465624   -0.659685    1.04316    0.305888   -25.7486      -15.5313
      9 -0.749915   -0.151375   -0.742794   -0.0452924    20.3753       -4.0625
```

```
[ ]: from sklearn.metrics import r2_score, mean_absolute_error
      print(r2_score(df_test['target'].values, df_test['prediction'].values))
      print(mean_absolute_error(df_test['target'].values, df_test['prediction'].
      ↪ values))
```

```
0.5920774016319219
15.96033360781515
```

```
[ ]:
```