

# Machine Learning Using Dask

# Machine Learning Using Dask

- DaskML provides scalable ML algorithms

# Machine Learning Using Dask

- DaskML provides scalable ML algorithms
- Works well on a single machine

# Machine Learning Using Dask

- DaskML provides scalable ML algorithms
- Works well on a single machine
- Can also be scaled up to a cluster of machines

# Single Machine Parallelism with sklearn

- Sklearn can use all cores on a single machine
- Manually setting hyperparameter  $n\_jobs = -1$

# Single Machine Parallelism with sklearn

- Sklearn can use all cores on a single machine
- Manually setting hyperparameter  $n\_jobs = -1$



# Machine Learning Using Dask



# Machine Learning Using Dask





# Dask on Cluster of Machines

```
from dask.distributed import Client
import joblib

client = Client() # Connect to a Dask Cluster

with joblib.parallel_backend('dask'):
    # Your normal scikit-learn code here
```

# Dask on Cluster of Machines

```
from dask.distributed import Client
import joblib

client = Client() # Connect to a Dask Cluster

with joblib.parallel_backend('dask'):
    # Your normal scikit-learn code here
```

# Dask on Cluster of Machines

```
from dask.distributed import Client
import joblib

client = Client() # Connect to a Dask Cluster

with joblib.parallel_backend('dask'):
    # Your normal scikit-learn code here
```

# Dask on a Single Machine

```
from dask_ml.preprocessing import Categorizer, DummyEncoder
from dask_ml.linear_model import LogisticRegression

lr = LogisticRegression()
lr.fit(data, labels)
```

# Machine Learning Using Dask

Install Dask ML

```
pip install dask-ml
```