In [1]:
```
from tpot import TPOTClassifier
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split

digits = load_digits()
X_train, X_test, y_train, y_test = train_test_split(digits.data, digit
                                                    train_size=0.75, t
```

```
/cluster/home/lucaschn/.local/lib64/python3.6/site-packages/sklearn/
ensemble/weight_boosting.py:29: DeprecationWarning: numpy.core.umath
_tests is an internal NumPy module and should not be imported. It wi
ll be removed in a future NumPy release.
  from numpy.core.umath_tests import inner1d

Generation 1 - Current best internal CV score: 0.9696105754417846
Generation 2 - Current best internal CV score: 0.9696105754417846
Generation 3 - Current best internal CV score: 0.9696105754417846
Generation 4 - Current best internal CV score: 0.9725767186833434
Generation 5 - Current best internal CV score: 0.9733228265344319

Best pipeline: ExtraTreesClassifier(input_matrix, bootstrap=False, c
riterion=entropy, max_features=0.6500000000000001, min_samples_leaf=
1, min_samples_split=5, n_estimators=100)
0.9733333333333334
```

In [5]:
```
%%time
pipeline_optimizer = TPOTClassifier(generations=5, population_size=20,
                                    random_state=42, verbosity=2)
pipeline_optimizer.fit(X_train, y_train)
```

```
Generation 1 - Current best internal CV score: 0.9696105754417846
Generation 2 - Current best internal CV score: 0.9696105754417846
Generation 3 - Current best internal CV score: 0.9696105754417846
Generation 4 - Current best internal CV score: 0.9725767186833434
Generation 5 - Current best internal CV score: 0.9733228265344319

Best pipeline: ExtraTreesClassifier(input_matrix, bootstrap=False, c
riterion=entropy, max_features=0.6500000000000001, min_samples_leaf=
1, min_samples_split=5, n_estimators=100)
CPU times: user 4min 43s, sys: 8.06 s, total: 4min 51s
Wall time: 4min 20s
```

In [6]:
```
%%time
pipeline_optimizer = TPOTClassifier(generations=5, population_size=20,
                                    random_state=42, verbosity=2, n_jo
pipeline_optimizer.fit(X_train, y_train)
```

```
Generation 1 - Current best internal CV score: 0.9696105754417846
Generation 2 - Current best internal CV score: 0.9696105754417846
Generation 3 - Current best internal CV score: 0.9696105754417846
Generation 4 - Current best internal CV score: 0.9725767186833434
Generation 5 - Current best internal CV score: 0.9733228265344319

Best pipeline: ExtraTreesClassifier(input_matrix, bootstrap=False, c
riterion=entropy, max_features=0.6500000000000001, min_samples_leaf=
1, min_samples_split=5, n_estimators=100)
CPU times: user 20.1 s, sys: 5.2 s, total: 25.3 s
Wall time: 1min 9s
```

In [4]:
```
from dask.distributed import Client
client = Client()
client
```

```
/cluster/home/lucaschn/.local/lib64/python3.6/site-packages/distribu
ted/dashboard/core.py:72: UserWarning:
Port 8787 is already in use.
Perhaps you already have a cluster running?
Hosting the diagnostics dashboard on a random port instead.
  warnings.warn("\n" + msg)
```

Out[4]:

Client
Scheduler: tcp://127.0.0.1:38251
Dashboard: http://127.0.0.1:45166/status (http://127.0.0.1:45166/status)

Cluster
Workers: 8
Cores: 48
Memory: 540.94 GB

In [7]:
```
%%time
pipeline_optimizer = TPOTClassifier(generations=5, population_size=20,
                                    random_state=42, verbosity=2, n_jo
pipeline_optimizer.fit(X_train, y_train)
```

```
Generation 1 - Current best internal CV score: 0.9674215243827646
Generation 2 - Current best internal CV score: 0.9769948190006887
Generation 3 - Current best internal CV score: 0.9822134722049327
Generation 4 - Current best internal CV score: 0.98369527891148
Generation 5 - Current best internal CV score: 0.98369527891148

Best pipeline: LinearSVC(PolynomialFeatures(BernoulliNB(input_matrix
, alpha=0.1, fit_prior=False), degree=2, include_bias=False, interac
tion_only=False), C=1.0, dual=False, loss=squared_hinge, penalty=l2,
tol=0.001)
CPU times: user 1min 45s, sys: 15.9 s, total: 2min 1s
Wall time: 3min 28s
```