In [1]:

```
import numpy as np
import pyarrow.feather as feather
import pandas as pd
```

In [2]:

```
import pyarrow.feather as feather
import pandas as pd
from surprise.model_selection import train_test_split, GridSearchCV, RandomizedSearc
from surprise import Dataset, Reader, BaselineOnly, accuracy
from joblib import parallel_backend
from surprise import SVD
```

In [3]:

```
df_train = feather.read_feather('netflix-5k.train.feather')
df_val = feather.read_feather('netflix-5k.validation.feather')
df_titles = feather.read_feather('netflix-5k.movie_titles.feather')
reader = Reader(rating_scale=(0, 5))
data = Dataset.load_from_df(df_train[['userID','movieID',
                                      'rating']],reader)
datav = Dataset.load_from_df(df_val[['userID','movieID',
                                     'rating']],reader)
trainset = data.build_full_trainset()
NA,valset = train_test_split(datav, test_size=1.0)
```

In [4]:

```
param_grid = {'n_factors': [125,150,175,200,225],
              'n_epochs':[40, 60, 80, 90, 110],
              'biased': [True, False],
              'init_std_dev': [0.2, 0.4, 0.6],
              'lr_all': [0.005, 0.0025, 0.00125],
              'reg_all': [0.04, 0.045, 0.05, 0.055],
              'init_mean': [0.5, 0.25, 0.125]
             }
with parallel_backend('multiprocessing', n_jobs=-1):
    gs = RandomizedSearchCV(SVD, param_grid, measures=['rmse'], cv=5, n_iter=8,
                            n_jobs=-1, joblib_verbose=10)
    gs.fit(data)

print(gs.best_score['rmse'])
print(gs.best_params['rmse'])

rmse_params = gs.best_params['rmse']
```

```
[Parallel(n_jobs=-1)]: Using backend MultiprocessingBackend with 8 con
current workers.
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:  1.1min
[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:  2.1min
[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  3.2min
[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  5.5min
[Parallel(n_jobs=-1)]: Done  30 out of  40 | elapsed:  6.8min remainin
g:  2.3min
[Parallel(n_jobs=-1)]: Done  35 out of  40 | elapsed:  7.2min remainin
g:  1.0min
[Parallel(n_jobs=-1)]: Done  40 out of  40 | elapsed:  7.9min remainin
g:    0.0s
[Parallel(n_jobs=-1)]: Done  40 out of  40 | elapsed:  7.9min finished

0.8446673153520337
{'n_factors': 225, 'n_epochs': 110, 'biased': False, 'init_std_dev':
0.2, 'lr_all': 0.005, 'reg_all': 0.055, 'init_mean': 0.25}
```

In [10]:

```
pd.DataFrame.from_dict([gs.best_params['rmse']])
```

Out[10]:

	n_factors	n_epochs	biased	init_std_dev	lr_all	reg_all	init_mean
0	225	110	False	0.2	0.005	0.055	0.25