AWS-foryou

examples

Example 1

Running sklearndiabetes.py as the user's algorithm.

In [1]:

```
import boto3
import numpy as np
import os
import pandas as pd
from sklearn import datasets
from sklearn import preprocessing
from sklearn.decomposition import PCA
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
from sklearn.svm import SVR
import time

os.chdir("..")

from awsforyou import aws_foryou
import examples.sklearn_diabetes as sdb
```

```
Using TensorFlow backend.
```

In [2]:

```
# writing the csv files for x and y (data_loc and target_loc respectively)
# datasize is 3 times larger than the standard diabetes dataset
x,y = sdb.get_diabetes(3)
```

In [3]:

```
# assign the locations of csv files
data_loc = './examples/x_diabetes.csv'
target_loc = './examples/y_diabetes.csv'
```

In [4]:

```
# reference start time
start = time.time()
best_model = sdb.run_sklearn_diabetes(data_loc, target_loc)
finish = time.time()
runtime = finish - start
print("runtime is %f seconds" % runtime)
print("best model is \n %s" % best_model)
```

```
linear regression score = 0.199718
best hyperparameters estimate from grid search =
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='sigmoid', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.213365
begining 6-components PCA decomposition
percentage of variance explained = 0.775656
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='auto',
  kernel='sigmoid', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.224709
runtime is 223.354946 seconds
best model is
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='auto',
  kernel='sigmoid', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
```

In [5]:

```
df = aws_foryou.aws_foryou("run_sklearn_diabetes(data_loc='examples/x_diabetes.csv', ta
rget_loc='examples/y_diabetes.csv')", 'examples.sklearn_diabetes')
df
```

```
linear regression score = -0.092879
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='auto',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.053999
begining 6-components PCA decomposition
percentage of variance explained = 0.831216
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=0.1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='aut
o',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.048403
point 1, iteration 1 complete.
linear regression score = -0.092879
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='auto',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.053999
begining 6-components PCA decomposition
percentage of variance explained = 0.831216
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=0.1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='aut
o',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.048403
point 1, iteration 2 complete.
linear regression score = -0.092879
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='auto',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.053999
begining 6-components PCA decomposition
percentage of variance explained = 0.831216
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=0.1, cache_size=200, coef0=0.0, degree=5, epsilon=0.1, gamma='aut
o',
  kernel='poly', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = -0.048403
point 1, iteration 3 complete.
linear regression score = 0.164878
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.267202
begining 6-components PCA decomposition
percentage of variance explained = 0.801720
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.252620
point 2, iteration 1 complete.
linear regression score = 0.164878
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.267202
begining 6-components PCA decomposition
```

```
percentage of variance explained = 0.801720
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.252620
point 2, iteration 2 complete.
linear regression score = 0.164878
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.267202
begining 6-components PCA decomposition
percentage of variance explained = 0.801720
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=3, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.252620
point 2, iteration 3 complete.
linear regression score = 0.101181
best hyperparameters estimate from grid search =
 SVR(C=2, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.157328
begining 6-components PCA decomposition
percentage of variance explained = 0.788614
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.162661
point 3, iteration 1 complete.
linear regression score = 0.101181
best hyperparameters estimate from grid search =
 SVR(C=2, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.157328
begining 6-components PCA decomposition
percentage of variance explained = 0.788614
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.162661
point 3, iteration 2 complete.
linear regression score = 0.101181
best hyperparameters estimate from grid search =
 SVR(C=2, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.157328
begining 6-components PCA decomposition
percentage of variance explained = 0.788614
repeat grid search with PCA-transformed data
best hyperparameters estimate from grid search =
 SVR(C=1, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
  kernel='linear', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
score from using best hyperparameters = 0.162661
point 3, iteration 3 complete.
Removing points data files.
WARNING:tensorflow:From /home/nawats/.local/lib/python3.6/site-packages/te
```

```
nsorflow/python/framework/op_def_library.py:263: colocate_with (from tenso
rflow.python.framework.ops) is deprecated and will be removed in a future
version.
Instructions for updating:
Colocations handled automatically by placer.
WARNING:tensorflow:From /home/nawats/.local/lib/python3.6/site-packages/te
nsorflow/python/ops/math_ops.py:3066: to_int32 (from tensorflow.python.op
s.math_ops) is deprecated and will be removed in a future version.
Instructions for updating:
Use tf.cast instead.
mnist runtime: 43.891830
```

The command above runs the algorithm "run_sklearn_diabetes" form examples.sklearn_diabetes through aws_foryou to produce a html report file.

In [ ]: