The Engineering World #DataScience 26

May 31, 2018

AKKAL BAHADUR BIST
DATA SCIENTIST AT
KATHMANDU INSTITUTE OF APPLIED SCIENCES (KIAS)
Center for Conservation Biology (CCB)

1 K-NEAREST NEIGHBOR CLASSIFICATION

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pylab import rcParams

import scipy
import urllib
import sklearn

from sklearn.neighbors import KNeighborsClassifier
from sklearn import neighbors
from sklearn import preprocessing
from sklearn.cross_validation import train_test_split
from sklearn import metrics
```

In [ ]:
```
np.set_printoptions(precision=4, suppress=True)
%matplotlib inline
rcParams['figure.figsize'] = 7, 4
plt.style.use('seaborn-whitegrid')
```

1.0.1 Setting your data into test and traines datasets

In [ ]:
```
address = 'mtcars.csv'
cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am
X_prime = cars.ix[:,(1,3,4,5)].values
Y = cars.ix[:,(9)].values
```

In [ ]:
```
X = preprocessing.scale(X_prime)
```

In [ ]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = .33, random_state =
```

1.0.2 Building and training your model with training data

```
In [ ]: clf = neighbors.KNeighborsClassifier()
        clf.fit(X_train, Y_train)
        print(clf)
```

1.0.3 Evoluting your model's prodiction against the test dataset

```
In [ ]: y_expect = Y_test
        y_pred = clf.predict(X_test)
        print(metrics.classification_report(y_expect,y_pred))
```

```
In [ ]: address = 'Advertising2.csv'
        cars = pd.read_csv(address)
        plt.plot(cars)
```

```
In [ ]: print(cars)
```

```
In [ ]: plt.hist(cars)
```

```
In [ ]: plt.bar(cars)
```