MACHINE LEARNING 7, 8, 9, 10 & 11

Assignment 1

MACHINE LEARNING 7 to 11: Assignment 1

Table of Contents

1.Introduction

This assignment will help you to consolidate the concepts learnt in the session.

2.1. Problem Statement: Machine Learning 7

In this assignment, students will be using the K-nearest neighbors algorithm to predict how many points NBA players scored in the 2013-2014 season.

A look at the data

Before we dive into the algorithm, let's take a look at our data. Each row in the data contains information on how a player performed in the 2013-2014 NBA season.

Download 'nba_2013.csv' file from this link:

https://www.dropbox.com/s/b3nv38jjo5dxcl6/nba_2013.csv?dl=0

Here are some selected columns from the data:

player - name of the player

pos - the position of the player

g - number of games the player was in

gs - number of games the player started

pts - total points the player scored

There are many more columns in the data, mostly containing information about average player game performance over the course of the season. See this site for an explanation of the rest of them.

We can read our dataset in and figure out which columns are present:

```
import pandas

with open("nba_2013.csv", 'r') as csvfile:

 nba = pandas.read_csv(csvfile)
```

2.2. Problem Statement: Machine Learning 8

In this assignment students have to find the frequency of words in a webpage. User can use urllib and BeautifulSoup to extract text from webpage.

Hint:

```
from bs4 import BeautifulSoup

import urllib.request

import nltk



response = urllib.request.urlopen('http://php.net/')

html = response.read()

soup = BeautifulSoup(html,"html5lib")
```

2.3. Problem Statement: Machine Learning 9

In this assignment students have to compress racoon grey scale image into 5 clusters. In the end, visualize both raw and compressed image and look for quality difference.

The raw image is available in spicy.misc package with the name face.

Hint:

```
import numpy as np
```

```
from sklearn import cluster, datasets

from scipy import misc
```

2.4. Problem Statement: Machine Learning 10

In this assignment students have to transform iris data into 3 dimensions and plot a 3d chart with transformed dimensions and colour each data point with specific class.

Hint:

```
import numpy as np

import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import Axes3D

from sklearn import decomposition

from sklearn import datasets
```

2.5. Problem Statement: Machine Learning 11

In this assignment students have to make ARIMA model over shampoo sales data and check the MSE between predicted and actual value.

Student can download data in .csv format from the following link:

https://datamarket.com/data/set/22r0/sales-of-shampoo-over-a-three-year-period#!ds =22r0&display=line

Hint:

Following is the command import packages

and data from pandas import read_csv

from pandas import datetime

from matplotlib import pyplot

from statsmodels.tsa.arima_model

import ARIMA from sklearn.metrics

import mean_squared_error def

parser(x):

return datetime.strptime('190'+x, '%Y-%m')

series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser)

NOTE:ThesolutionsharedthroughGithubshouldcontainthesource code usedand thescreenshotoftheoutput.

3.Output

This assignment consists of 1000 marks and needs to be submitted in GitHub. You can follow GitHub submission guide provided to do the same.