import matplotlib.pyplot as plt

import pandas as pd

%matplotlib inline

import numpy as np

from pandas import read\_csv

path = r"C:\pima-indians-diabetes.csv"

headernames = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']

data = read\_csv(path, names = headernames)

array = data.values

X = array[:,0:8]

Y = array[:,8]

data.shape

(768, 9)

data.head()

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Preg** | **Plas** | **Pres** | **Skin** | **Test** | **Mass** | **Pedi** | **Age** | **Class** |
| 0 | 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 2 | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 3 | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 4 | 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |

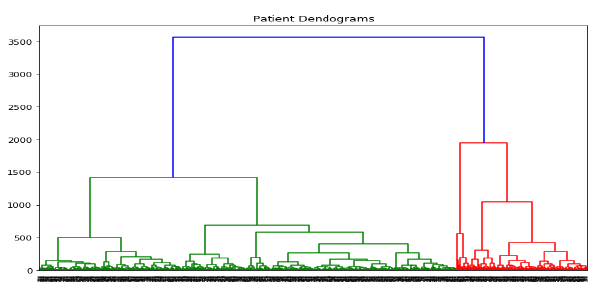
patient\_data = data.iloc[:, 3:5].values

import scipy.cluster.hierarchy as shc

plt.figure(figsize = (10, 7))

plt.title("Patient Dendograms")

dend = shc.dendrogram(shc.linkage(data, method = 'ward'))



from sklearn.cluster import AgglomerativeClustering

cluster = AgglomerativeClustering(n\_clusters = 4,

affinity = 'euclidean', linkage = 'ward')

cluster.fit\_predict(patient\_data)

plt.figure(figsize = (10, 7))

plt.scatter(patient\_data[:,0], patient\_data[:,1],

c = cluster.labels\_, cmap = 'rainbow')

