PHYS 3142 Spring 2021
Computational Methods in Physics
Assignment 8
Due: 11:59 p.m. 11th Apr. 2021

Before you submit your assignment, do remember:

1. the due day
2. submit a report which contains your figures and results along with your code
3. make sure your code can run
4. do not forget to write comments in your codes.
5. label your figures and describe your results

The basic scoring rubric is :

1. **If you submit the assignment after the deadline or do not submit the report, you can only get up to 80% of grade**
2. **If there is any kind of plagiarism, all of the student involving will get zero mark! (except that the one can really prove the code is written by himself or herself and others copied it without telling him or her)**

1. Principal Component Analysis(PCA) in Iris Data_set

In this question, we're going to apply the PCA method to the Iris dataset, where there are 150 data points in total, 50 for Setosa, 50 for Versicolour and 50 for Virginica. The data set contains 4 kinds of information of the plants: sepal length in cm, sepal width in cm, petal length in cm and petal width in cm. Now consider import the Iris plants dataset from sklearn.datasets, and we're going to reduce the dimension and visulize the data! (You can find more information about the data set: `https://scikit-learn.org/stable/datasets/toy_dataset.html`)

1. First, load your data_set and have a look at the form that is given to you:

```
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
import pandas as pd
import numpy as np


dataset = load_iris()
X, y = load_iris(return_X_y = True)
df = pd.DataFrame(X, columns=dataset.feature_names)
N = 50
```

```
In [857]: df
Out[857]:
     sepal length (cm)  sepal width (cm)  petal length (cm)  petal width
(cm)
0                  5.1               3.5                1.4
0.2
1                  4.9               3.0                1.4
0.2
2                  4.7               3.2                1.3
0.2
3                  4.6               3.1                1.5
0.2
4                  5.0               3.6                1.4
0.2
```

X returns you the matrix as shown, while y gives you the label for different species of plants. (0 for Setosa, 1 for Versicolour, 2 for Virginica)

Please use plt.subplot to obtain 4×4 scatter plots for all the four features, i.e. plot data for 2 features every time. Please use different colors for different species and correctly label the points and axes.

2. In order to visualize the data, the highest dimension that we could make a plot with respect to is three. Pick the first three, (or any three as you like) and plot a 3D scatter graph. Again please use different colors and label the axes. Can you see the connection between the 3D plot and the 4×4 subplots? (If you forget how to do the 3D plots, please refer to Lec2, Page 84)

3. Now we want to visualize the data with 4 features.(4 dimensions) We could deal with this problem using PCA. (You need to define your own function to find the covariance matrix) Please show the 2d scatter plot for the 'PCAed' matrix. Is this graph capable of demonstrating the variance for species as expected?

The following questions are optional, just for exercise and will not be marked.

2. PCA for the Ising Model

Consider again the Ising Model for the ferromagnetic system, please use MCMC (Markov chain Monte Carlo) method to simulate the Ising model in square lattice with 10×10 sites. Please use periodic boundary conditions and choose the total number of points to be 1000. The Hamiltonian is

$$H = -J \sum_{<i,j>}^{N} S_i S_j$$

where J is set to be 1. Now we obtain the 1000×100 matrix consisting of 1 and -1, which has 25 dimensions of 'features'.(We assume that the direction of spins on each site is independent) We take the points at temperature equals to 1.6, 2.8 and 3.7 as 3 classes of points. Please use PCA method to reduce the dimensions and show the final scatter plot. Remember to use different colors for different groups and correctly label them.

$$X = \begin{pmatrix} \uparrow & \downarrow & \uparrow & \cdots & \uparrow & \uparrow & \uparrow \\ & & & \vdots & & & \\ \downarrow & \uparrow & \downarrow & \cdots & \uparrow & \downarrow & \uparrow \end{pmatrix}_{M \times N},$$

In our case, M=1000, and N =100