

IITR-TSW PG CERTIFICATION PROGRAM 2024

Module: Advanced Machine Learning

Assignment

1) Generate the 2D data using the following piece of code:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_moons
Xmoon, ymoon = make_moons(300, noise=.08, random_state=0)
plt.scatter(Xmoon[:, 0], Xmoon[:, 1]);
```

- i) Build a clustering instance using Gaussian Mixture Model to cluster the above data in two clusters.
- ii) Visualize the result of above clustering. Explain the reason for clustering error.
- iii) Use GMM as a generative model for generating 100 new such sample points.
- iv) Again perform the clustering of those 100 points using model described in part (a).
- v) **Optional:** generate some face images from LFW dataset using the GMM based generative model.

2) Download the Liver patient data from the following sources:

<https://www.kaggle.com/uciml/indian-liver-patient-records>

Use the following 07 features from this dataset:

Age; Total_Bilirubin; Direct_Bilirubin; Alkaline_Phosphatase;
Alamine_Aminotransferase; Total_Protiens; Albumin

Your task is to predict whether a patient suffers from a liver disease using above features. Split your data into test and train.

First use a random forest algorithm for performing this task.

Then, use a Adaboost Classifier to perform similar task.

Compare the accuracy of these two algorithms.

-End of the Assignment-