**PGP in AI/ML**

**Classification - Assignment 1**

**Submission Date: 2359 hrs on 24-08-2019**

**Total Marks: 25**

The following questions are to enhance your understanding of basic concepts and definitions. You are expected to answer these questions on your own without referring to internet.

1. List the limitations of some prediction measures (like Accuracy). Provide the scenarios highlighting the limitations **[6M**]
2. Suggest measures countering these shortfalls. **[4M]**
3. For the dataset -2 (liver disease dataset uploaded along with the assignment), implement kNN classifier using Python. Find out the optimal k for the case study and put up appropriate reasoning/arguments for choosing the optimal k. **[15M**]

Please follow the following steps.

1. Import the libraries [1M]
2. Load the dataset [1M]
3. Remove/replace missing values (if any) [2M]
4. Split features and labels [1M]
5. Split train and test data [1M}
6. Calibrate optimal k using kNN [2M]
7. Calculate accuracy measures [2M]
8. Cross validate to better the accuracy [3M]
9. Plot ROC curves to compare different models built in the process [2M]

Submission Details

Text answers – id\_classification\_assignment1.doc

Code - id\_knn.ipynb

Contact Details

(You should put up queries in the discussion forum of the corresponding assignment folder only)

1. Himanshu - himanshuchawla437@gmail.com
2. Mohit - mohitrathore15@gmail.com
3. Parthasarathy - pdparthasarathy.03@gmail.com