**CS697A – Topic in Computer Science – Machine Learning**

**Summer 2021**

**Assignment 1**

Due date : July 11, 2021 at 11:00pm

Instructor : Zehra Cataltepe

**PURPOSE:**

To understand how to use various models in machine learning to find the confidence and error. And knowing which model works best.

**WHERE TO SUBMIT ASSIGNMENTS**

Please submit through the class Blackboard site. Please zip and upload all your files using filename studentID\_HW1.zip. Submit a zip file of the Jupyter Python notebook as specified below. Download the jupyter notebook named hwk1 and edit the contents on this file and rename it to your studentID\_Hwk1.ipynb.

Include your name, ID, your groupID and your groupmates’ names and IDs in all files. Each group member must submit the same file.

Fill in expected row and submit as a separate excel file, studentID\_HW1\_expected.xlsx in your zipped file.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Question | Q1 | Q2 | Q3 | Q4 | TOTAL |
| Max | 4 | 2 | 2 | 2 | 10 |
| Expected |  |  |  |  |  |

**POLICY:**

Collaboration in the form of discussions is acceptable, but you should write your

own answer/code by yourself. Cheating is highly discouraged for it could mean a zero or negative grade from the homework. If a question is not clear, please let me know (via email, during office hour or in class).

**DATA:**

Download the dataset from:

<https://www.kaggle.com/uciml/pima-indians-diabetes-database/data>

Questions [Total 10pts]:

**Q1 [4pts]:** Use naive Bayes, logistic regression and 3-nn classifiers (library) to train on the training sets and compute training and validation errors for each fold. The target label is Outcome. You can use a library to do the partitioning or you can write your own code.

**Q2 [2pts]:** Is the error of naive bayes <0.2 with confidence 0.9

**Q3 [2pts]:** Do naive bayes and knn have the same error?

**Q4 [2pts]:** Do the three classifiers have different errors?