*Introduction to Machine Learning (BRI-507)*

**Midterm Project**

Submission type: Individual

Deadline: May 9th, 2020 (end-of-the-day)

**This midterm project replaces midterm exam.**

1. ***Task Description:*** Given a train dataset “train\_data.csv”, build your model for the below regression and classification tasks
   1. **Input Features**: morphological phenotypes from a brain MR image
      1. Colum E-BV (ST\*\*\*CV) : Cortical Volume of 70 brain regions, i.e., 70-dimension
      2. Colum BW-EN (ST\*\*\*TA) : Average Thickness of 70 brain regions, i.e., 70-dimension
      3. Note that there are missing values.
      4. If necessary, do your own preprocessing for data clearning.
   2. **Task 1 (3-class Classification)**: Predict the diagnosis group of subjects
      1. Colum A (DX\_bl): Diagnosis group of subject  
         0: Cognitive normal (CN)

1: Mild cognitive impairment (MCI)  
2: Alzheimer’s disease (AD)

* 1. **Task 2 (3-logit Regression)**: Predict the cognitive assessment scores of subjects in the test dataset of “sample\_evaluation\_data.csv”.
     1. Colum B (ADAS11): Alzheimer's Disease Assessment Scale (11 questions version)   
        [range: 0~70]
     2. Colum C (ADAS13): Alzheimer's Disease Assessment Scale (13 questions version)  
        [range: 0~85]
     3. Colum D (MMSE): Mini-Mental State Examination  
        [range: 0~30]

1. ***Project Requirement***
   1. Build your best model based on the given train dataset “train\_data.csv”
   2. Report scores for 10-fold cross validation
   3. Compare with at least two other models
      1. i.e. Total of at least 6 models should be built  
         (One best model and two comparing models for each task)
   4. Analyze and discuss your models and results in markdown cells
   5. TAs will assess your best model with a test dataset, and grade your project accordingly
2. ***Implementation Requirement***
   1. Use the Google Colab (<https://colab.research.google.com/>)
   2. You may use any library of your choice (e.g. Scikit-learn)
   3. (**IMPORTANT**) **Submit two .ipynb files and model weight files**
      1. In each of the .ipynb file, write your name, ID, and the link to your Colab project
      2. First .ipynb file: “midterm\_project\_yourID.ipynb”
         1. This should contain your code, model comparisons, analysis, discussion *etc.*
      3. Second .ipynb file: “midterm\_evaluation\_yourID.ipynb”
         1. This should contain code for loading and evaluating your best model
         2. Follow further instructions in “midterm\_evaluation\_template.ipynb

<https://colab.research.google.com/drive/18h_fhRih_PnQXtsm7jA2ZvaEqln8aVrv>

1. **Grading Factors**
   1. Evaluation scores on test dataset
      1. Ground truth of the test dataset will not open to students.
      2. Scores will be ranked in terms of AUC and MAE scores, and used as a reference for grading.
   2. Comparison, analysis, discussion *etc.*
      1. Write in English
      2. Be thorough and precise

TA’s note: Beware of plagiarism (we use proprietary plagiarism checkers in addition to BlackBoard plagiarism checker). Ask the TAs if you have any questions.