Handwritten Character Prediction Competition

For the handwritten character prediction competition, your task is to predict the lowercase letters a-z from their binary images. The training and test data are provided as sequences of images where each sequence is extracted from a handwritten word.

The competition is worth 10% of the course grade. You should participate **individually** in the competition.

1 How To Participate

1.1 Accessing Competition Page

The competition has been hosted on Kaggle and can be accessed through following url : `https://kaggle.com/join/cs5339predictioncompetition`. The training and test data for the competition can be accessed through the "Data" tab on left.

1.2 Software

You may use any software to do the learning and predictions. Scikit Learn has many algorithms. For deep learning, you can consider packages such as Apache Singa, Tensorflow, Theano, etc. You may also write your own learning and prediction programs.

1.3 Making a Submission

After you have trained your models using training data, you can make a submission on Kaggle through "Make a Submission" tab on left. You can find details about submission file format in the "Evaluation" section under "Information" section of left. A sample submission file has also been provided in the data section for your help.

Before you make a submission make sure that your team name has following format (Your name)-(Your student id). This is required for correct grading. Team name can be set by going to the "My Team" tab on the left.

1.4 Leaderboard

Public leaderboard will show the percentage of correct predictions on part of the test set. You can see the rank of your trained model through it. The rankings on Private Leaderboard might be slightly different as it ranks the performance of your predictions on a different part of the test set.

2 Grading

The competition closes at midnight on Sunday 19 March. You will be evaluated based on the percentage of correct predictions of the letters on the Private Leaderboard at the end of the competition. Following will be the grading criteria:

A: accuracy $>= 90\%$
B: $80\% <=$ accuracy $< 90\%$
C: $70\% <=$ accuracy $< 80\%$
D: $60\% <=$ accuracy $< 70\%$
E: $50\% <=$ accuracy $< 60\%$
F: accuracy $< 50\%$

In addition, you are required to submit your source code to the IVLE workbin. Your assignment is considered incomplete if the source code is not submitted.

3 How to do Well

Understand the data.

- What properties does the data have?
- How do you exploit the properties with the learning algorithm?