IITM-CS4011 : Principles of Machine Learning — Release date: 16 Oct, 2017, 11:55 am

Contest — Deadline: 12 Nov, 2017, 11:55 pm

- The goal of this contest is to test your overall understanding of the concepts.
- Collaborations and discussions with other groups are strictly prohibited.
- You are allowed to use any language and tool.
- You have to turn in a detailed report mentioning every step used for training the model with hyper-parameter specification. Explain all the pre-processing, processing and post-processing approaches that were tried. We need to see justifications for why certain techniques were used, some unique insights that you've gained, your explanations to why certain approaches worked and others didn't, etc.
- Typeset your report in LaTeX. Note that **the report is an important part of contest. Ensure that it is as complete and detailed as possible.**
- Submit the well documented codes along with the report.
- Check Moodle regularly for updates regarding the contest.
- **Please start early to utilize the number of allowed submissions effectively.**

Registration Instructions

- Form a team of 2 and fill this form *once*.
- Both the team members should create their respective accounts on Kaggle and register in this competition as a single team. Both of you will be allowed to make submissions. The submissions, however, will be counted together for the team.
- Please do not participate in the competition without a partner. Find a partner, form a team and then enter the contest. You may not be allowed to form a team later.

Dataset details and Evaluation

- This is a classification problem of 29 classes with 9501 training instances of 2600 dimensions. The evaluation metric is the mean F1-score.
- Train_data and their respective classes can be found in Contest_Data. Test_data without class labels is also given.
- You are free to use any technique to train your model(s).

- During the leaderboard phase, you can upload your labels for the given test data and get the f1-score of prediction for 30% of the test_data. The ranking and performance for each team will be displayed on the same set.
- The final ranking will be done on the private test_data.
- The number of submissions for a team is restricted to 5 per day.
- You will be allowed to select 2 of your submissions for the final submission.
- The baseline model is Logistic Regression. It obtains ~ 0.17 mean F1-score on the public test dataset. Submissions below this baseline will not be considered.