IITM-CS5011 : Introduction to Machine Learning — Given on: Oct 22, 12 am

Contest — Due on : Nov 15 , 11:55pm

The goal of this contest is to test your understanding of the concepts.

This has to be done in a group of 3. Collaborations and discussions with other groups are strictly prohibited.

You are allowed to use any language or tool.

You have to turn in detailed report mentioning the step by step procedure for training the model with complete specification of hyperparameter used. Explain all the approaches tried.

Typeset your report in Latex. Note that Report is an important part of contest so it should be as complete as possible.

If you have any codes then submit the well documented codes.

You have to check the Moodle discussion forum regularly for updates regarding the contest.

- This is a classification problem of 12 classes with 30,000 training instances of 3702 dimensions.
- Train data and their respective classes can be found in DATASET. Test data without class labels is also given.
- Train your model to get best accuracy.
- You are free to use any technique to train the model.
- The model will be evaluated on the basis of accuracy.
- A different set of test data will be used for evaluation so avoid over-fitting on train data.
- The website will be set up soon and the leaderborad phase will start. In the leaderboard phase you can upload your labels for the given test data and get the accuracy of prediction. This can be used to compare performance with other teams.
- Final evaluation will be done on a different set of test data which will also include the already given test data.

  Note: This set will be released in the last week of contest.
- Baseline model is Tensorflow 0-hidden layer model with adam optimization. It gets 34.5333% accuracy on test set after training on 1000 mini batches of size 64 each. Accuracy on test set falls to roughly 32% after training for 100,000 mini batches.
- Teams that get accuracy below the baseline (of 34%) will not be evaluated.