Topic Covered: Logistic regression and regularisation
Instructions to the students:
Questions:

The CSV files belongs to the famous titanic dataset which contains the details related to survived and non-survived passengers.

Develop logistic regresssion model (without regularisation) for this classification problem.

- Convert all the given data (you may also drop some features if you feel, it wont contribute to the present problem) to the numerical and/or categorical values.
- Develop the logistic regression model (without regularisation) with train dataset given in train.csv file. Obtain the error in the training dataset.
- With the developed model, predict the values for test dataset given in test.csv and calculate the error in the prediction.

Develop the logistic regresssion model (with regularisation) for this classification problem.

- Divide the test dataset into cross validation dataset and test dataset
- Develop the logistic regression model (with regularisation) with train dataset given in train.csv file. Obtain the error in the training dataset and cross validation dataset.
- To choose best value of regularisation parameter, plot λ vs $J_{CV}(\theta)$ and λ vs $J_{train}(\theta)$ and choose the best value of λ
- With the developed model, predict the values for test dataset given in test.csv and calculate the error in the prediction.

Compare the above two models and ellobrate your inferences.

For any clarifications: Reach me at **harimurugan@nitj.ac.in**