

Assignment - Deep Feedforward Neural Networks

The goal is to predict whether a passenger survived or not using the Kaggle Titanic Challenge, based on attributes such as age, gender, passenger class, where they boarded and so on.

First, log into Kaggle and go to the [Titanic Challenge](#) to download 'train.csv' and 'test.csv'. The data is already divided into a training set and a test set. However, the test data does not contain the labels: your goal is to train the best possible model using the training data, so make your predictions on the test data and load it into Kaggle to see your final score.

First, assess the training set attributes. Some attributes contain missing data. This indicates that it may be unnecessary to include them in the model.

Then evaluate the performance of a feedforward neural network (MLP) on this dataset.

To improve this result, you can:

- Adjust hyperparameters (number of layers and neurons in each layer) using cross validation and grid search;
- Pre-process the features, for example:
 - replace SibSp and Parch by their sum;
 - try to identify parts of names that correlate well with the Survived attribute (e.g. if the name contains "Countess" then survival seems more likely),
 - Try converting numerical attributes to categorical attributes: for example, different age groups had very different survival rates. So it might help to create an age category and use that instead of age. Likewise, it can be useful to have a special category for people traveling alone, as only 30% of them survived.

You must send a report containing the learning curves and confusion matrix of the initial model and the improved models. Please, also include the architecture and parameters used in the improved models, as well as an analysis of the results obtained and the impacts related to the variation in network architecture/parameters.