Beihang University
School of General Engineering

Beijing, 15th, June
Wei Lin and Yida Ding

Data Management and Artificial Intelligence Homework 15

Task 1 ***Surviving chance of Titanic passengers*** *(25 minutes)*

The well-known perishing Titanic sank in 1912 killed 1502 out of 2224 passengers and crew. In this task you are expected to find out if some people were more likely to survive (e.g. women and children) according to the passenger information data. The following subtasks are given as follow:

(1) Load file **titanic_dataset.csv** using pandas. The label information is stored in the first column *'survived'*. As name (column *"name"*) and ticket (column *"ticket"*) information may not affect the analysis, you need to remove both columns and change type of the column *"sex"* from char to integer (0, 1) as non numerical data can not be handled.

(2) Establish the network structure for neural network. *tflearn.input_data, tflearn.fully_connected* functions are recommended to establish the input layer, hidden layer and output layer. Implement optimizer (such as "*adam*") and loss function (such as "*categorical_crossentropy*") through *tflearn.regression*.

Hint: Two hidden layers with 32 neurons are preferred to achieve better performance and the final activation function in output layer can be "*softmax*".

(3) Meanwhile, construct the whole model with *tflearn.DNN* and train the model with *.fit* function. Batch size can be set to 16 to reduce memory usage and 200 epoch will train a good enough model for prediction. Use the model to predict the surviving chance of two main characters (**Jack and Rose**) by inputting:

dicaprio = [3, 'Jack', 'male', 19, 0, 0, 'N/A', 5.0000]

winslet = [1, 'rose', 'female', 17, 1, 2, 'N/A', 100.0000]