# Assignment 3: Machine Learning

Submission: Wednesday April 27th
Groups of maximum 2 students

Prof. Fabio A. González
Intelligent Systems - 2016-I
Maestría en Ing. de Sistemas y Computación

---

1. Follow the Kaggle tutorial Titanic: Machine Learning from Disaster. In particular the sections:
   (a) Getting Started With Python.
   (b) Getting Started With Python II.
   (c) Getting Started With Random Forests.
2. Train a Support Vector Machine (SVM).
   (a) Ensure that the features are numerical and normalized.
   (b) Use a RBF kernel. Explain what a RBF kernel is.
   (c) Use 10-fold cross validation to find a good value for the hyper-parameters $C$ (regularization coefficient) and $\gamma$ (RBF shape parameter). Use a logarithmic scale for the values of the parameters (eg. $2^i$ for $-5 \leq i \leq 10$).
   (d) Draw a plot with the performance for the different combination of hyper-parameter values.
   (e) Report the results on the training dataset using a confusion matrix.
   (f) Submit the results to Kaggle and include a screenshot of your score and position.
3. Model comparison.
   (a) Split the data in a training (70%) and test (30%) subsets.
   (b) Train a SVM, a Naïve Bayes model, and a random forest on the test subset.
   (c) Explain what a ROC curve is.
   (d) Compare the performance of the models using confusion matrices. Discuss the results.
   (e) Compare the performance of the models using ROC curves. Discuss the results.

The assignment must be submitted as a Python notebook through the following Dropbox file request, before midnight of the deadline date. The file must be named as `is-assign3-unalusername1-unalusername2.ipynb`, where `unalusername` is the user name assigned by the university (include the usernames of all the members of the group).