

Homework 1

Many details of this assignment are left intentionally vague. The goal of this assignment is to give you a sense of the challenges a machine learning practitioner typically faces. No knowledge of machine learning is required to complete this assignment. However at least basic knowledge of Python is needed.

Please download the Titanic dataset from Kaggle.com and study it:

<https://www.kaggle.com/c/titanic>

Your goal is to develop an algorithm to make a prediction for each passenger whether that passenger will survive the sinking. Use your knowledge of Python and NumPy. One possible direction would be to write some code to load the dataset into a NumPy array and try to understand which attributes (columns) can help predict survival.

Now implement your algorithm in Python. Evaluate how accurate it is.

I do not expect you to solve this problem. However, you must give it a try even if it is just to understand how difficult this problem is.

Please submit the assignment using Sakai. The information about what to submit is included in a separate document (see my Logistics/ folder).

Good luck!