

## Assignment 1: Feature Selection using searching techniques

Due on: 10-FEB-2018 (11:59 PM)

Max Marks: 50

In machine learning, the fundamental problem is to learn (or approximate) a function  $f$  that maps the input  $X = \{x_1, x_2, \dots, x_M\}$  to an output  $Y$ . The function approximation process uses prior available data points  $\{X_i, Y_i\}$ ,  $i = 1, \dots, n$  where  $X_i$ 's are vectors of real numbers and  $Y_i$ 's are real numbers. Typically, the complete set of input features  $\{x_1, x_2, \dots, x_M\}$  is not useful and a subset of them  $\{x_1, x_2, \dots, x_m\}$ ,  $m < M$  is used for approximation. Selection of subset of features is known as Feature selection or variable selection. There are three broad categories of feature selection methods, filter methods, wrapper methods, and embedded methods. This assignment will focus on wrapper methods as these methods use searching techniques to find the required subset of features.

In this assignment each group has to select two searching techniques as part of wrapper method and implement it in Python. The code should be well commented and should be written in such a way that it can be tested with any data set. A write-up (not more than three pages long) is also required to be submitted that will describe the representation of states, the search strategy, and any other important points that you think is relevant to the assignment.

### What to submit?

1. Source code in Python +ReadMe file (35 Marks)
2. Report (not more than 3 pages) (15 Marks)

### Recommended Readings:

1. Blum and P. Langley, **Selection of relevant features and examples in machine learning**, Artificial Intelligence, 97(1-2):245-271, 1997.
2. Guyon and A. Elisseeff, **An Introduction to Variable and Feature Selection**, Journal of Machine Learning Research, 3, 1157-1182, 2003.
3. G. Chandrashekar and F. Sahin, **A Survey on Feature Selection Methods**, Computers and Electrical Engineering, 40, 16-28, 2014.