**REPORT ON FEATURE SELECTION**

Abstract:

In machine learning feature selection is the process of selecting a subset of relevant features. The main idea behind it is that the data contains many features that are redundant or irrelevant and thus can be removed without causing any loss to our data and thus reducing the computational time and complexity. Various techniques are used for the method of feature selection.

We have used two wrapper method algorithms namely *Hill Climbing* and *Genetic algorithm*.

**Pre-processing of Dataset:**

We have pre-processed the dataset in the following way:

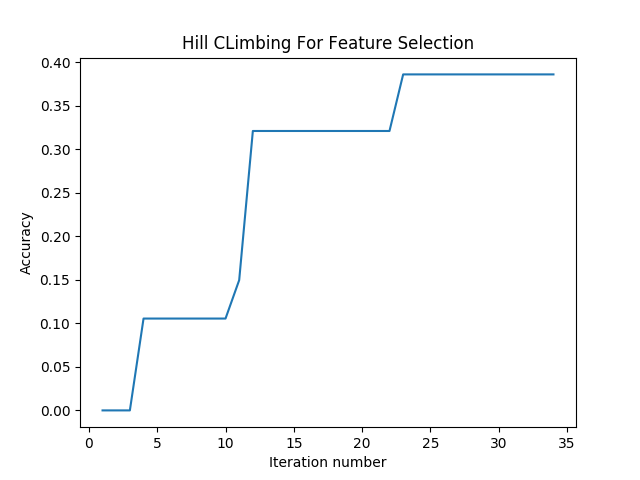
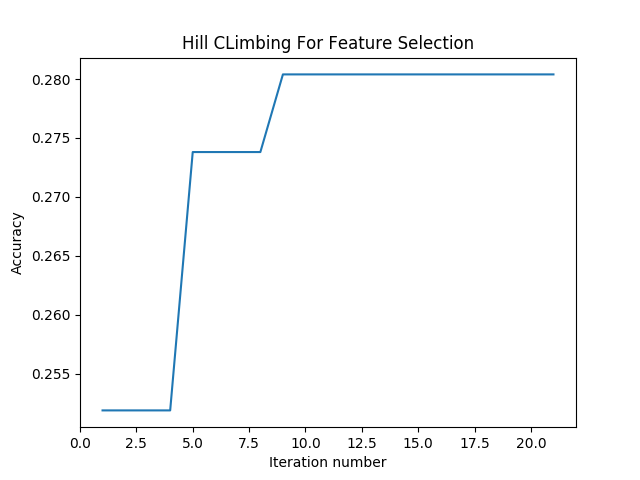
1. Columns having Null Values and ‘?’ are dropped.
2. Strings have been converted to integer values using the ASCII values of characters.

**Hill Climbing:**

It is heuristic search based algorithm used to determine an optimum value in a local space. It is an iterative algorithm that starts with an arbitrary solution and try to finds a better solution by incrementally changing a single element of the solution.

Our implementation:

1. Initial state: array of length equal to number of features with all elements initialised to 0
2. Create successor states by flipping 1 bit in the current state in each step and find out accuracy of that state and the state with the best accuracy is selected as next state
3. Continue creating successor states and flipping bits one at a time till we find a state with all successors having worse accuracy than this state.

Regression Classification

**Genetic Algorithm:**

It is a metaheuristic based algorithm inspired by the natural process of evolution. It is used to generate high quality solutions by using bio inspired operators like mutation, crossover and selection.

Our implementation:

1. Initially we start with a population which consists of 10 states created randomly
2. Then continue the following till number of generations < 30:
3. Loop through the population size 10 and select 2 individuals as parents randomly with tournament selection method.
4. Randomly Crossover the 2 parents to create a new child and mutate the child with a small probability and evolve the population.
5. We calculate the fitness of each state by passing the data set to the Support vector regression library.
6. We’ll calculate the fittest states among them.
7. We’ll then consider the states with fittest values and will select the features which are active (i.e. 1) among these states.