Introducing multi-objective GP

CSCI6506 Sandbox 3(b)

Assignment due date: 10th Feburary, 2015

Task definition

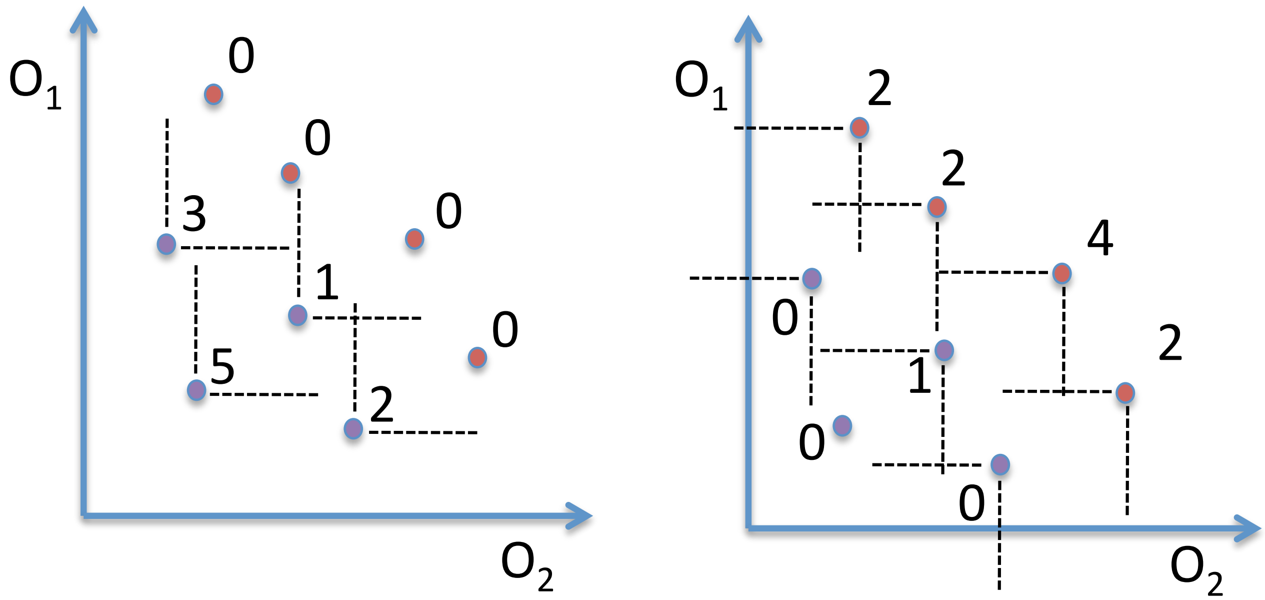
- Your goal is to extend the framework you developed for classification in Sandbox 2 to use a Pareto formulation for the objective.
  - The two objectives you will assume take the form of the Accuracy and Class-wise detection rate from Sandbox 2.
  - Also assume that the Data Subset is configured to compose samples with each class represented equally.
- Pareto dominance is defined as follows [3]:[1]

  A solution a_i is said to dominate another solution a_j, if both of the following conditions are true:[2]

  1. Solution a_i is no worse than a_j in all objectives, or $f_k(a_i) \geq f_k(a_j); \forall k$; where f_k is the kth objective (fitness) function and objectives are to be maximized.
  2. Solution a_i is better than a_j in at least one objective, or $f_k(a_i) > f_k(a_j); \exists k$; where f_k is the kth objective (fitness) function and objectives are to be maximized.
- The definition for Pareto dominance (i.e., equivalence) alone, however, does not represent the entire story. Given that we require a scalar ranking of a set of individuals a mechanism for interpreting the outcome of the Pareto dominance criteria is necessary. You will consider the impact of assuming two different Pareto-based dominance measures:
  1. **Dominance rank:** is the count of the number of *other solutions* that dominate each candidate solution (Figure 1.(a)). This tends to result in a broad front of solutions being maintained.
  2. **Dominance count:** is the count of the number of solutions that *each candidate solution* dominates (Figure 1.(b)). This tends to result in a concentration of solutions about some focal point.

[1] See [1] summary of the development of the evolutionary multi-objective topic.

[2] The widely used compact definition has the form: $a_i \succ a_j \leftrightarrow \forall k[f_k(a_i) \geq f_k(a_j)] \wedge \exists m[f_m(a_i) > f_m(a_j)]$



(a) Dominance ranking

(b) Dominance count

Figure 1: Quantifying Pareto dominance through: (a) dominance ranking – better has a lower rank; and, (b) dominance counting – better has a higher count. Objective maximization assumed.

- The next issue that we need to deal with is:
  - who will get to be parents and,
  - who are replaced at each generation.

  In particular there can be a significant problem with replacement error.[3]

- Given that we are not in a position to identify an ordering of the phenotypes, we will therefore limit **variation operators** to mutation alone (there can be more than one form of mutation).

- We will assume a $(\mu + \lambda)$ formulation for the **selection operators** in which all members of population $P(t)$ generate one offspring, i.e. $\lambda = \mu = |P|$. Both parents and offspring are ranked using Pareto dominance, and the top $|P|$ individuals retained as the basis for population $P(t+1)$.

- The next issue with Pareto formulations is that diversity maintenance is not straight forward (see above footnote regarding support for the crossover operator).
  - Identify 5 diversity mechanisms that have been proposed in the literature for use with multi-objective formulations of GP and include within your report (these should be examples from the published literature).

[3]A solution associated with one peek in the search space is replaced by the offspring associated with a completely different peek, potentially resulting in the loss of an entire mode of the search space.

- One final mechanism for incorporating multiple objectives in evolutionary computation is to introduce periodic switches in the objective [2]. Thus, rather than attempt to support multiple objectives simultaneously, there is only one objective. However, every 5 or 10 generations the objective is switched. Return to your solution for Sandbox 2, this time with periodic switching between the Accuracy and Class-wise DR metric.

Reporting

- Write a short 2 page summary answering the following questions:
  - Given that your result from training is a 'front' of non-dominated solutions, how do you go about selecting your 'champion' solution for deploying during test?
  - What were the 5 diversity mechanisms you managed to identify in the literature?
  - How effectively does the dominance rank based fitness versus dominance count based fitness promote the exploration of the objective space? [Hint: you might want to visualize the candidate solutions from the last generation in objective space.]
  - What 'hidden penalties' do Pareto formulations encounter when employed with genetic programming?
  - How effectively does the objective switching formulation for fitness promote the exploration of the objective space?
  - Did the multi-objective formulations compare with your scalar solutions from Sandbox 2?
- Your report should be emailed to `mheywood@cs.dal.ca` before midnight on the sandbox due date.

References

[1] C. A. Coello Coello. Evolutionary multi-objective optimization: A historical view of the field. *IEEE Computational Intelligence Magazine*, pages 28–36, Feb 2006.

[2] S. Doncieux and J.-B. Mouret. Behavioral diversity with multiple behavioural distances. In *IEEE Congress on Evolutionary Computation*, pages 1427–1434, 2013.

[3] A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Natural Computing Series. Springer, 2003.