. **Homework 10: Efficient Global Optimization and Statistical Comparison， SOP&NFL, Performance Profile**

**Due: Wednesday,** **April 18, 2018**

1. **Surrogate Optimization with Pareto Selection (SOP) – Center Selection (15 points)**

Recall the “center selection” methodology discussed in class for the SOP algorithm. Assume that you have to select three points as center points from twelve evaluated points in an SOP iteration (the actual problem is a minimization problem). Table 1 provides a sorted list of the evaluated points according to 1) Their non-domination rank (best rank to worst rank) and 2) Best to worst objective function value on each rank.

Table 1: Evaluated points sorted by 1) Non-Domination Rank 2) Best to worst objective function value, for SOP center selection

|  |  |
| --- | --- |
| **Evaluated Point No.** | **F(x) - Objective Function Value** |
| 10 | 1.38 |
| 6 | 24.94 |
| 5 | 62.06 |
| 2 | 3.69 |
| 1 | 7.52 |
| 4 | 16.38 |
| 12 | 44.39 |
| 3 | 71.90 |
| 7 | 9.30 |
| 9 | 46.97 |
| 8 | 83.69 |
| 11 | 51.29 |

* 1. **(5 points)** Given the sorted points in Table 1 and their corresponding objective function values, can you deduce the number of non-dominated fronts? If the answer is yes, report the number of non-dominated fronts.
  2. **(5 points)** Table 2 provides a 12-by-12 grid / matrix where each cell, (i,j), within the grid corresponds to the distance between evaluated point i and evaluated point j (Hence, the grid is and should be symmetric). Assume that **T = {6,4,7}**, where the set T corresponds to the set of evaluated points which are Tabu from center selection in the current SOP iteration (e.g, evaluated point 6 is Tabu). Furthermore, assume that **r** is the neighborhood search radius for all evaluated points (hint: radius rule), and **r = 1**. Given the T, r and the information provided in Tables 1 and 2, identify the three points that will be selected as centers in the current SOP iteration.
  3. **(5 points)** Now assume that **T = {5,2} r = 1**. Identify the three centers that will now be selected as centers in the current SOP iteration. (all other information remains as it is).

Table 2: Matrix where each cell(i,j) depicts the distance (in decision space) between evaluated points i and j

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Distance(i,j)** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **1** | 0 | 3.70 | 3.22 | 0.99 | 8.93 | 7.14 | 0.20 | 0.43 | 7.86 | 0.75 | 3.65 | 8.18 |
| **2** | 3.70 | 0 | 0.48 | 2.71 | 5.23 | 3.44 | 3.89 | 3.27 | 4.16 | 2.95 | 0.04 | 4.48 |
| **3** | 3.22 | 0.48 | 0 | 2.23 | 5.71 | 3.92 | 3.41 | 2.79 | 4.65 | 2.47 | 0.44 | 4.97 |
| **4** | 0.99 | 2.71 | 2.23 | 0 | 7.94 | 6.15 | 1.18 | 0.56 | 6.87 | 0.24 | 2.66 | 7.19 |
| **5** | 8.93 | 5.23 | 5.71 | 7.94 | 0 | 1.79 | 9.12 | 8.50 | 1.07 | 8.18 | 5.27 | 0.74 |
| **6** | 7.14 | 3.44 | 3.92 | 6.15 | 1.79 | 0 | 7.33 | 6.71 | 0.72 | 6.39 | 3.48 | 1.04 |
| **7** | 0.20 | 3.89 | 3.41 | 1.18 | 9.12 | 7.33 | 0 | 0.62 | 8.06 | 0.94 | 3.85 | 8.38 |
| **8** | 0.43 | 3.27 | 2.79 | 0.56 | 8.50 | 6.71 | 0.62 | 0 | 7.44 | 0.32 | 3.23 | 7.76 |
| **9** | 7.86 | 4.16 | 4.65 | 6.87 | 1.07 | 0.72 | 8.06 | 7.44 | 0 | 7.11 | 4.21 | 0.32 |
| **10** | 0.75 | 2.95 | 2.47 | 0.24 | 8.18 | 6.39 | 0.94 | 0.32 | 7.11 | 0 | 2.91 | 7.44 |
| **11** | 3.65 | 0.04 | 0.44 | 2.66 | 5.27 | 3.48 | 3.85 | 3.23 | 4.21 | 2.91 | 0 | 4.53 |
| **12** | 8.18 | 4.48 | 4.97 | 7.19 | 0.74 | 1.04 | 8.38 | 7.76 | 0.32 | 7.44 | 4.53 | 0 |

1. **No Free Lunch Theorem (10 points)** (This question uses notation given in lecture slides)

Assume you want to compare a genetic algorithm with an evolutionary search (ES)

algorithm by how well they can maximize a function. The decision vector is a binary

string of length 5. The value of the objective function is an integer between 1 and 10. So

the domain is the set of binary strings of length 5 and the range is {1,2,…,10}. The initial

population for the optimization is picked randomly (uniform distribution) from the

domain. All comparisons below assume algorithms are run so that 1000 objective function evaluations of J are made at distinctly different binary strings for each J.

a) **(3 points)** How many possible problems (i.e. **how many different *JK ’s***) are there for

this situation?

b) **(3 points)** How many of the possible *JK ’s* have an objective function value of strictly

less than 6 for all possible binary strings of length 5? To show you understand this concept, give one example of one objective function *JK(x)* such that *JK (x)* is less than 6 for all *x* in domain *S* (*This is very easy—the question is* *just to make you think concretely about what a J in a subset of Z is).* Let us call this set *G={ JK , K=1,…,H| JK(x)<5}*. So the question is how big is *H=|G|*? What is the **ratio of |Z| to |G|** (expressed as a power of 2)? Approximate that as function of 10 (e.g. ratio |Z|/|G|= where b= is an integer). This base 10 question is to make sure you appreciate the difference in size of these function sets

c) **(4 points)** By the No Free Lunch (NFL) Theorem we can show that:

Where P(M|Jk ,n,GA) / P(M|Jk,n,ES) is the probability (for GA and ES, respectively) that in the n iterations, the best solution found for Jk is M. Assume that you divide the *JK ’s* in Z into two disjoint sets Z1 and Z2 so that |Z1| + |Z2|= |Z|, and the intersection of Z1 and Z2 is empty. Assume also that ¾ of the J’s are in set Z1 and the remainder are in set Z2. You want to compare GA with Evolutionary Search (ES). Assume you can estimate that the average (over all *JK* in *Z1*) probability is .5 that GA will find M to be the best solution in 1000 evaluations of each *JK(x)* and that the probability is .4 that ES will find M to be

the best solution in 1000 evaluations of each *JK(x).* Then what is the **difference** between the average probability for GA and average probability for ES of findingthe M to be the best solution for each of all the *JK (x)* that are in Z2? **Give an equation and explain your answer**. If you use an equation summing over theJ’s be sure to specify from which set the sum is over.

1. **Efficient Global Optimization**
2. Assume you want to minimize an objective function. You have fitted a Gaussian Process model from a set of initial design points and you get the current optimal solution . Now you want select the next evaluation point from 8 candidate points by evaluating the expected improvement function. The predictive mean and variance of the candidate points are given as

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 0.2345 | -5.360 | 24.55 |
| 0.3541 | 0.2415 | 6.57 |
| 0.4256 | 2.618 | 6.54 |
| 0.6112 | 0.726 | 13.73 |
| 0.7535 | -4.542 | 29.94 |
| 0.8451 | -9.243 | 16.41 |
| 0.9365 | 8.882 | 9.97 |
| 0.9842 | -3.2454 | 3.24 |

Which point will be selected for next evaluation?

1. Explain the advantages and disadvantages of EGO.
2. **Leave-one-out cross-validation (LOOCV)**

|  |  |
| --- | --- |
|  |  |
| -0.1 | 0 |
| 0.7 | 1 |
| 1.0 | 1 |
| 1.6 | 0 |
| 2.0 | 1 |
| 2.5 | 1 |
| 3.2 | 0 |
| 3.5 | 0 |
| 4.1 | 1 |
| 4.9 | 1 |

1. Consider the following dataset with one real-valued input and one binary output. The following question assume that for any input , you predict by taking the same value as its nearest neighbor. What is the leave-one-out cross validation error? Give your answer as the number of wrong predictions.
2. Consider the following dataset with real-valued input and output. You want to validate your Gaussian Process model by LOOCV. The cross-validated prediction for each point is denoted as , and the cross-validated variance is denoted as .

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 0.2345 | -9.7856 | -5.360 | 24.55 |
| 0.4256 | 5.7465 | 2.618 | 6.54 |
| 0.6112 | 1.874 | 0.726 | 13.73 |
| 0.7535 | -11.3968 | -4.542 | 29.94 |
| 0.9365 | 9.4067 | 8.882 | 9.97 |

* 1. How to evaluate whether the model is valid at a single point given the cross-validated prediction and variance? Why?
  2. What is the LOOCV error? Give your answer as the number of points that fails the cross validation.
  3. If there exists a point that fails the cross validation, what are the possible strategies you can do to improve the model fit?

1. **(No code) Statistical Comparisons: The table below shows the objective function value for the best solution in each trial for three different algorithms applied to the same problem:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Trial** | **SA** | **GA** | **GS** |
| 1 | 91.94 | 147.90 | 47.66 |
| 2 | 77.13 | 97.88 | 150.53 |
| 3 | 10.93 | 39.76 | 97.04 |
| 4 | 18.6 | 204.48 | 82.62 |
| 5 | 28.63 | 488.83 | 99.89 |
| 6 | 86.52 | 113.00 | 76.52 |
| 7 | 64.58 | 141.97 | 87.84 |
| 8 | 22.23 | 53.76 | 51.73 |
| 9 | 59.75 | 408.20 | 147.51 |
| 10 | 134.11 | 226.95 | 115.98 |
| Mean | 59.44 | 192.27 | 95.73 |
| Std. Dev | 39.52 | 148.35 | 34.90 |

(i). Make a boxplot for the data provided above (Use the Matlab command BOXPLOT). Comment on your plot: How do the means and variances compare for each of the algorithms? Are there any outliers? Which algorithm performed the best in your opinion and why?

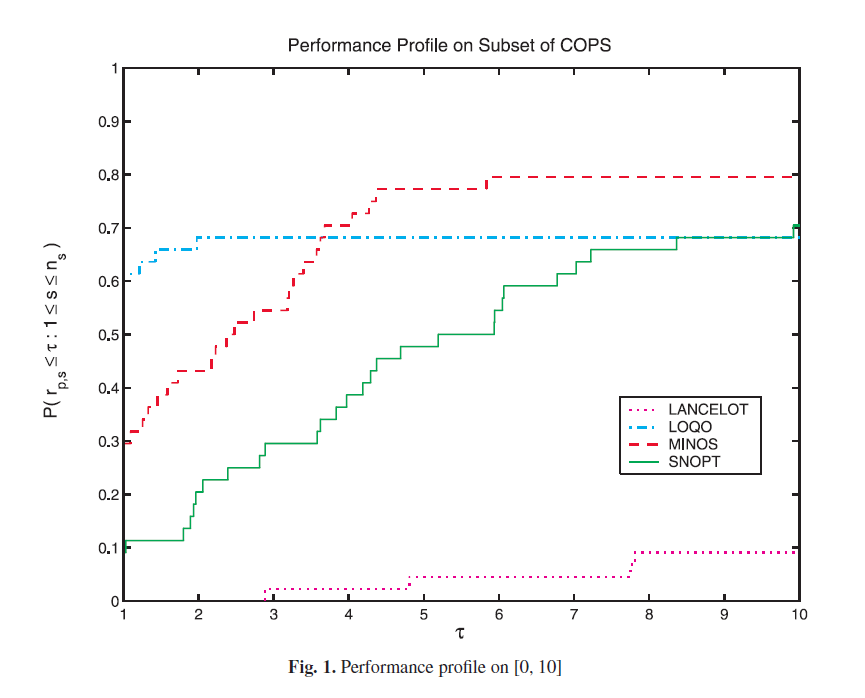
(ii). Plot empirical CDF’s for the data above using the plotting position formula provided in class (i/[n+1]). Comment on your plot: Which algorithm appears to perform the best? Is there any evidence of stochastic dominance?

(iii). Perform all pairwise comparisons (three in all) of mean objective function value of best solution using a two-sample t-test. State your hypothesis. Report your test statistic and p-values for each comparison. At = 0.05 what is your conclusion for each test?

(iv). If you were told that SA and GS had the same starting solution in each trial would you perform a different test for comparing these two algorithms? Explain why or why not. If you decide to perform another test, state your hypothesis and report its p-value and your conclusions at = 0.05. Compare your test results to those in (iii).

(v). Perform non-parametric comparisons for all pairwise tests performed in (iii). Report your test statistic and p-values for each comparison. At = 0.05 what is your conclusion for each test? Compare your test results to those in (iii).

**6. (No code) Performance Profile. Firgure 1 is Performance Profile about 4 algorithms tested on COPS test problems set. X-axis is tolerance level ranges from 1-10, Which measures the allowed ratio between the best solver with the current solver. Y-axis is the proportion of problems that a given solver can solve.**

****

Please compare each algorithm based on the Performance Profile above.