Exam 1

Due: Tuesday, November 5, 2019

1. Using the baseball data and keeping the total number of candidates examined close to 2,500, change the steepest ascent local search algorithm to employ 2-neighborhoods.
   (a) Include the function in your printed version of the exam.
   (b) If you had R random starts each of S steps searching a neighborhood of size k of the p parameters, then what is the total number candidates examined?
   (c) How many restarts and steps did you uses? In total how many candidates did you examine?
   (d) What was the smallest AIC you obtained?
   (e) Include a plot similar to the one on page 10 of the Combinatorial Optimization notes.

2. Modify the genetic search algorithm to search for K clusters which minimize the total within-group sum of squares of p characteristics. The total within-group sum of squares is

$$TWGSS = \sum_{k=1}^{K} \sum_{i \in C_k} \sum_{j=1}^{p} (x_{ij} - \overline{x}_{kj})^2$$

   where $C_K = \{i : \text{individual } i \text{ is in cluster } k\}$, x_{ij} is value of characteristic j measure in individual i, $\overline{x}_{kj}$ is the average value of characteristic j in cluster k.
   (a) Include the function in your printed version of the exam.
   (b) Use your function to group the wines from problem 3.8 into 2, 3, and 4 clusters based on the 13 chemical characteristics. What were the total within-group sum of squares you found for 2, 3, and 4 clusters.

3. Let p be an interpolating polynomial of order m of the function f at nodes $a = x_0 < \ldots < x_m = b$. That is

$$p(x) = \sum_{i=0}^{m} f(x_i) \prod_{j=0, j \neq i}^{m} \frac{(x - x_j)}{(x_i - x_j)}.$$

   (a) Show that

$$f(x) - p(x) = \frac{f^{m+1}(\xi_x)}{(m+1)!} \prod_{i=0}^{m} (x - x_i)$$

   for some $\xi_x \in [a, b]$.
   Note:

$$F(y) = f(y) - p(y) - \frac{f(x) - p(x)}{\prod_{i=0}^{m} (x - x_i)} \prod_{i=0}^{m} (y - x_i)$$

   will be useful.
   (b) Use 3a) to show that for $m = 1$

$$\int_a^b f(x) - p(x) dx = \frac{-(b-a)^3 f^{(2)}(\xi)}{12}$$

   for some $\xi \in [a, b]$.