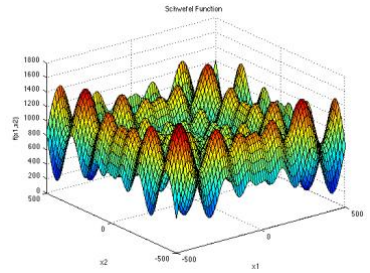
MSSE 277B: Machine Learning Algorithms

Homework assignment #2: Simulated Annealing
Assigned Feb. 2 and Due Feb. 14

1. (9pt) Classical simulated annealing. We will use the Schwefel function for D=10 in order to find its global minimum using CSA



$$f(x_1, x_2 \dots x_D) = 418.9829\text{xD} - \sum_i^D x_i \left(\sin\left(\sqrt{x_i}\right)\right)$$

$$x_i \in [-500,500] \text{ for } i = 1, \dots, D$$

In which we use the visitation function of a random displacement along each dimension

$$x_i = x_i + (2 * URN - 1) \times \Delta, \text{ with } \Delta=0.5 \text{ for } i = 1, \dots, D$$

(a) (4pt) Fill in the blanks in the provided simulated annealing code. Use a linear ($T_{t+1}=T_t-\alpha$) cooling schedule with α=0.5, and initializing T_{SA}=3000K, to perform CSA until the temperature reaches 30K and 10K, and record the function values. How long is your cooling schedule? Check against the debugging outputs. Given the stochastic nature of CSA, it would be best to report at least 3 runs for each lower bound temperature. Do you find better solutions when cooling to the lower temperature?

(b) (2pt) Choose logarithmic cooling ($T_k=T_{SA}/(1+ T_{SA} \log(1+k)/3\sigma_{curr}$), where k is counter for number of cooling cycle) and σ_{curr} is an adjustable parameter, with two initial temperature T_{SA} = 3000K and 6000K. Use σ_{curr} = 1000 and k = 6000. Reconsider questions (a). Do these cooling schedules converge better than linear cooling?

(c) (3pt) Create your own annealing schedule (cooling and heating cycles) to see if you can find better solutions. Use a local optimization technique on your CSA answer, can you find even better solution?

2. (11 pt) Clustering and simulated annealing. Clustering is a widely used technique in exploratory data analysis that we will examine later using unsupervised learning for classification of objects into groups. But for now we will consider a popular meta-heuristic for solving it using CSA. In this case we would like to cluster *N* data points into *K* clusters by solving the minimization of the following cost function:

$$J(N,K) = \sum_{i=1}^{N}\sum_{j=1}^{K} w_{ij} d_{ij}^2$$

$$w_{ij} = \begin{cases} 1 & if\ point\ i\ is\ assigned\ to\ cluster\ j \\ 0 & otherwise \end{cases}, \quad 1 \le i \le N \quad and \quad 1 \le j \le K$$

where d_{ij} is the Euclidean distance between point i and the center of cluster j, and condition on w_{ij} ensures that a point is defined to be in one of the distinct clusters K.

In this problem you are given a data set of N =178 wines, all grown in the same region in Italy, but derived from K=3 different cultivars. You can find this dataset on *BCourse -> Files -> Datasets -> wines.csv*. A chemical analysis was used to determine the quantities of 13 chemical constituents found in each of the wines, and our job is to classify each wine into one of the three different cultivars based on those chemical descriptors. Use CSA to determine the most optimal clustering of the 178 wines into their most likely cultivar, given the following specifications

(a) (1pt) Normalize your chemical descriptor data for each attribute by subtracting off the mean and dividing by the standard deviation.

(b) (1pt) Given the initial categorization of the 178 wines into the 3 clusters according to *Start assignment* column in the dataset, determine the centroid of each of the three clusters. The centroid for this problem is a 13-D vector where each entry is the mean of a variable for the observations in that cluster.

(c) (1pt) Given the centroid, determine the value of the cost function for this initial categorization. Check against the debugging output.

(d) (4pt) Fill in the blanks in the provided simulated annealing code. Use CSA with a visitation function in which a randomly chosen wine i is moved from its present cluster j to another randomly chosen cluster $k \neq j$. One epoch corresponds to attempting to move all N wines between clusters, i.e. there are N Metropolis steps, at each temperature. Use a start temperature of 500, and use a geometric cooling schedule($T_{t+1}=\alpha T_t$) with $\alpha=0.999$ and total of 5000 steps, again using at least 3 runs of CSA. Check your final temperature against debugging output. Report all 3 solutions and the wine members as part of each cluster. Validate your result using the provided code. How well is the assignment?

(e) (4pt) Adapt your code in 2(d). Now use CSA with a visitation function in which a randomly chosen centroid j is updated as a random walk for each of its 13 components

$x_i = x_i + (2 * URN - 1) \times \Delta$; with Δ=0.01

In this case one epoch corresponds to moving all $K = 3$ cluster centers at each temperature, reassigning all wines to their nearest centroid, and evaluating the new cost function. Check against the debugging output to make sure you are assigning wines correctly. Use a start temperature of 500, and use a geometric cooling schedule with $\alpha=0.999$ and total of 5000 steps, again using at least 3 runs of CSA. Report all 3 solutions and the wine members as part of each cluster. Is this a better solution than found in (d)?