Engi 9867-Advanced Computing Concepts in Engineering
Winter 2019- Memorial University

Assignment 1:

Due date: 15th February 2019
Total Marks: 100

Problem 1: The following 10 points are given and we want to fit a conic with equation $ax^2 + bxy + cy^2 + dx + ey = 10000$ to them by minimizing mean square of error: $P_1 = (61, 35)$, $P_2 = (39, 37)$, $P_3 = (39, 44)$, $P_4 = (21, 25)$, $P_5 = (10, 17)$, $P_6 = (20, 9)$, $P_7 = (26, 16)$, $P_8 = (42, 12)$, $P_9 = (42, 28)$, and $P_10 = (50, 24)$.

(a) Find the parameters (a, b, c, d, e) using steepest descent algorithm with step size $\mu = 0.01$. Show the number of iterations required to reach a convergence and the final value of mean square of error (5 marks).
(b) Find the parameters (a, b, c, d, e) using genetic algorithm with maximum population size $N_{max} = 8$. Show the number of generations needed to reach a convergence and the final value of fitness function (7 marks).
(c) Assuming the initial population as $X_1 = (1.5, 0, 1.3, 0, 0)$, $X_2 = (2.5, -2, -1, 0, 2)$, and $X_3 = (0, 1.5, 0, -1.2, 2.8)$.
i) Evaluate the fitness for each individual and arrange them with the fittest first and the least fit last (2 marks).
ii) Perform the crossover between the two fittest individuals using one-point crossover at point c (3 marks).
iii) Perform the crossover between the first and third fittest individuals using two-point crossover at points b and d (3 marks).

Problem 2: The following 5 points are given: $P_1 = (4.25, 2.12)$, $P_2 = (3.9, 2.27)$, $P_3 = (1.97, 3.47)$, $P_4 = (4.41, 0.47)$, and $P_5 = (2.12, 3.66)$. Cluster these points into $k = 2$ clusters using:

(a) k-means algorithm (7 marks)
(b) single linkage HAC algorithm and sketch its dendrogram (8 marks)
(c) spectral clustering algorithm with $\sigma = 1$ (8 marks).

Problem 3: Consider data points generated from two different equi-probable classes. Class 1 has the distribution $p(x|c = 1)$ $N(0, 1)$ and class 2 has the distribution $p(x|c = -1)$ $N(2, 1)$. Compute the posterior probability $p(c = 1|x)$ (5 marks).

Table 1: A subset of IRIS dataset

Sample #	SL	SW	PL	PW	Specie
1	5.8	4.0	1.2	0.2	Setosa
2	5.7	4.4	1.5	0.4	Setosa
3	6.1	2.8	4.0	1.3	Versicolor
4	6.3	2.5	4.9	1.5	Versicolor
5	7.3	2.9	6.3	1.8	Virginica
6	6.2	2.8	4.8	1.1	Virginica

Problem 4: Implement Gaussian mixture model (GMM) for clustering and use it to cluster the Iris data into three clusters when each mixture includes g=2 Gaussian distributions(12 marks).

Problem 5: First randomly split Iris data into two equal-size subsets for training and testing and then classify the testing data using each of the following algorithms. For each case, calculate the classification error.

(a) KNN with k=1,5,10,15 (8 marks).
(b) SVM with linear Kernel (3 marks)
(c) SVM with polynomial kernel with degree 2 and 3 (6 marks)
(d) SVM with RBF kernel with σ=1 and 2 (6 marks).

Problem 6: Table 1 shows six samples of IRIS dataset. Use this data and:

(a) Develop a fully grown decision tree using CART algorithm and explain your solution (12 marks).
(b) Apply the first level post-pruning based on purity metric (5 marks).

Note: Except problems 1 and 5, for other problems you have to write your own implementation and are not allowed to use Matlab or Python functions.