

ECE/CS/ME 539 Introduction to Artificial Neural Networks and Fuzzy Systems

Homework #3

This homework covers the following topics: support vector machine, clustering (SOM, Gaussian mixture), radial basis network

1. (35 points) *support vector machine*

- (a) (6 points) Find the distance (positive or negative values) between a plane in a 3D space, described by the equation:

$$2x + 3y - z = 4.$$

and the following points $P1 = (0, 1, 3)$, $P2 = (-2, -1, 4)$, $P3 = (2, -2, 2)$. Are these points on the same side of this plane?

- (b) (4 points) Consider the set of training data:

Table 1. A training data file

<i>Feature</i>	<i>label</i>	<i>feature</i>	<i>label</i>
$x_1 = (0,0,0)$	$d_1 = -1$	$x_5 = (1,0,0)$	$d_5 = 1$
$x_2 = (0,0,1)$	$d_2 = 1$	$x_6 = (1,0,1)$	$d_6 = -1$
$x_3 = (0,1,0)$	$d_3 = 1$	$x_7 = (1,1,0)$	$d_7 = -1$
$x_4 = (0,1,1)$	$d_4 = -1$	$x_8 = (1,1,1)$	$d_8 = 1$

Find the corresponding dual cost function (as a function of α_i 's only) which has the general form below using the values specified in table 1.

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$$

- (c) (5 points) Consider a **second order** ($p=2$) polynomial kernel: $K(x,y) = (1+x^T y)^2$. Write a short program to compute the 8 by 8 kernel matrix \mathbf{K} , whose $(i, j)^{th}$ element is $K(x_i, x_j)$. What is the dimension of the corresponding transformed space $\phi(x)$? Submit the source code of the program as part of the answer.
- (d) (15 points, CC) Apply an implementation of the SVM classification algorithm to classify above training data. You may **modify** the code from svmdemo.m to suite your needs or use other existing svm packages. You need to report the following to receive credit for this part of question: (i) the package you used. (ii) the driver program you developed to use the package. (iii) the printout of the answer including $\{\alpha_i; 1 \leq i \leq 8\}$, and the support vectors. (*Hint*: You may have to try different types of kernels with different parameters. Polynomial kernel with order ≥ 3 and radial basis kernel are know to work).
- (e) (5 points) A hypothetical SVM model has the following values of α and support vectors:

α	<i>Support vector</i>	d
1	(1, -1, 1)	+1
0.5	(0, 2, -1)	-1
1	(-1, 0, 2)	-1

Suppose that the linear kernel is used. Compute the output of this SVM model when the input feature vector is (0.3, 0.8, 0.6).

2. (30 points) *Kmeans clustering algorithm*

- (a) (5 points) Given a set of cluster centers $\{w_i; 1 \leq i \leq c\}$. The data points x will be assigned to a cluster center w_i if $d(w_i, x) = \|w_i - x\|$ is minimum among all other cluster centers. This assignment will partition the feature space into c voronoi regions $\{V_i; 1 \leq i \leq c\}$ such that $V_i = \{x | \|w_i - x\| \leq \|w_k - x\|, k \neq i\}$. In a 2D feature space, if the cluster centers are located at $\{(-1, 0), (1, 1), (2, -3)\}$, draw these points and the boundary of corresponding voronoi regions within rectangular area having lower-left corner point $(-2, -4)$ and upper-right corner point $(3, 4)$.
- (b) (4 points) In the maximization step of the kmeans algorithm, we find new position of the cluster center, given a set of data points $\{x_i\}$ that are assigned to the same cluster. Define the within-cluster distortion as

$$D = \sum_{i=1}^{N_i} (x_i - w)^2$$

Find an expression of w that will minimize D . Show your derivation.

- (c) (6 points) To apply the kmeans algorithm, the number of the cluster center must be specified. One criterion is to use the *within-cluster scatter matrix* S_W and *between-cluster scatter matrix* S_B as defined in the class notes. Based on the definition of these matrices, derive the following equations:

a.
$$m = \frac{1}{N} \sum_{i=1}^c N_i m_i$$

b.
$$S_T = S_B + S_W.$$

Provide necessary details of your derivation.

- (d) (10 points, CC) Download data files `hw2p2a.txt` and `hw2p2b.txt`. Use kmeans clustering algorithm to cluster these data points. Use 2, 4 and 10 clusters for each data file. Plot the resulting clustering centers and voronoi regions similar to that shown in `clusterdemo.m`. You may copy and modify `clusterdemo.m` to solve this problem.
- (e) (5 points) Consider the clustering of the following eight points $(-1.5, 0)$, $(-0.5, 0)$, $(0.5, 0)$, $(1.5, 0)$, $(-1.5, 1.8)$, $(-0.5, 1.8)$, $(0.5, 1.8)$, $(1.5, 1.8)$ into two clusters. There are two possible choices of clustering centers:

Choice 1. $(0, 0)$, and $(0, 1.8)$

Choice 2. $(-1, 0.9)$, and $(1, 0.9)$

Which of these two choices has smaller distortion? (assume Euclidean distance measure is used. Run kmeans clustering algorithm on these data samples and verify your answer.

3. (10 points, CC) *Self Organizing Map*

Download data file `hw2p3.txt`. It contains 20 points within a square bounded by $(0, 0)$ and $(1, 1)$. Use a 1D array of 20 neurons, each labeled with a number 1 to 20.

- (a) (5 points, CC) Develop a SOM clustering program (you may **modify** from somdemo.m) using 10 neurons arranged in a linear array with each neuron labeled with 1 to 10. **Submit a copy of your program and a figure similar to that given in somdemo.m.**
- (b) (5 points, CC) Manually or using a program **to assign** each of the 20 points to a neuron (cluster center). Denote nc to be a 1 by 10 array such that $nc(i)$ = number of data points assigned to i -th neuron. **Measure the 2D distance between successive neurons.** Assign the first neuron with a coordinate of 0 in a 1-D number line. The coordinate of 2nd neuron equals to the distance from it to the first neuron. The coordinate of the 3rd neuron equal to the coordinate of the 2nd neuron plus the 2D distance from the 2nd neuron to the 3rd neuron, and so on. **Draw a bar graph** of nc versus each neuron's 1D coordinate just computed. This bar graph can be regarded as a 1D histogram approximating the 2D distribution of the given data points. Submit materials highlighted in bold face words.

4. (25 points) *radial basis network*

A radial basis function can be represented as:

$$F(x) = \sum_{i=1}^N w_i \phi(\|x - x_i\|)$$

where $\phi(r)$ is the radial basis function. Three examples of $\phi(r)$ are:

- (i) Multiquadrics: $\phi_m(r) = \sqrt{r^2 + c^2}$, $c > 0$,
- (ii) Inverse multiquadrics: $\phi_i(r) = 1/\sqrt{r^2 + c^2}$, $c > 0$, and
- (iii) Gaussian: $\phi_g(r) = \exp\{-r^2/(2\sigma^2)\}$, $\sigma > 0$,

Suppose that there are four pairs of training samples:

$$\{(x_i, d_i); 1 \leq i \leq 4\} = \{(-2, 1), (-1, -1), (0, -1), (2, 0.5)\}$$

Assume that $c = \sigma = 1$. By imposing the interpolation condition $F(x_i) = d_i$, a linear system of four equations

$$\Phi \mathbf{w} = \mathbf{d}$$

can be obtained where $\mathbf{w} = [w_1, w_2, w_3, w_4]^T$, and $\mathbf{d} = [1 \ -1 \ -1 \ 0.5]^T$.

- (a) (5 points) Derive the Φ matrix and find its eigenvalues for each of the three radial basis functions. Also, discuss your observations on these eigenvalues briefly.
- (b) (5 points) For each of the radial basis functions listed above, find the weight vector \mathbf{w} , and plot the approximated function $F(x)$ over the interval $[-3 \ 4]$ at increment of 0.05.
- (c) (15 points, CC) Down-load two files hw4p4r.txt (training file) and hw4p4t.txt (testing file). Use the training file to develop a RBF network and use the testing file to test the performance. Note the testing file can not be used during the development of the network. Perform the following tasks:
- (i) (5 points, CC) Develop a RBF type 1 network using Gaussian RBF with $\sigma = 0.5$. Choose at least three values of regularization parameter λ so that the training data is fitted best. Discuss how do you decide the fitting is good *without* using the testing file. **Submit** a) the source code, b) the model parameters (weights) developed, c) a table listing different values of λ and corresponding training errors. d) A plot of the

- model that fitting with training data using different values of λ of your choice (at least three should be shown). e) Another plot of fitting the model of your choice that fit the testing data.
- (ii) (5 points, CC) Develop a generalized RBF (type II) network using Gaussian RBF. Use two different the number of cluster centers J ($1 \leq J \leq 10$). For each value of J , **submit** a) compute the G matrix and the G_0 matrix, and the weights. b) the corresponding plots of the model fitting with the training data and the testing data.
- (iii) (5 points, CC) For RBF type I network with a Gaussian RBF function, discuss influence of different values of σ on the results in part (i). **Submit** a) a plot of fitting training data using at least three different values of σ while $\lambda = 0.1$.