

Name: _____

Final Exam

Statistical Pattern Recognition Department of Computer Engineering Sharif University of Technology Fall 2006

1. You must work out all the problems.
2. Show all your work in the space provided to justify your answers.
3. You may use your text book & class notes in this exam.

No	Max	Score
Problem 1	40	
Problem 2	20	
Problem 3	30	
Problem 4	15	
Total	105	

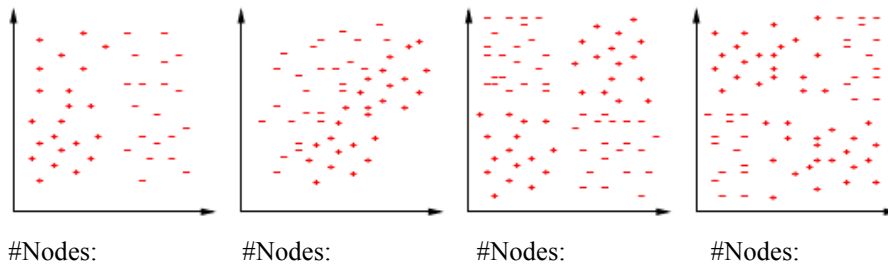
Good Luck!

1. Short Answer and Conceptual questions

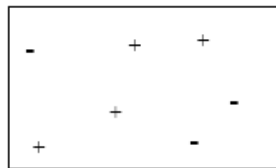
a. True/False

1. Every continuous function can be computed by some multi-layer perceptron with arbitrary unit function.
2. If we use mean-squared error to train a perceptron, we will always successfully separate linearly separable classes.
3. We would expect the support vectors to remain the same in general as we move from a linear kernel to higher order polynomial kernels.
4. Any decision boundary that we get from a generative model with class-conditional Gaussian distributions could in principle be reproduced with an SVM and a polynomial kernel of degree less than or equal to three.

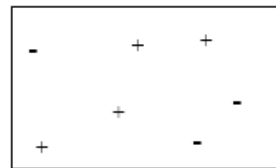
b. The Following diagrams represent training examples for a Boolean-valued concept function plotted in feature space. In each case how many nodes is needed for correctly classification by a perceptron?



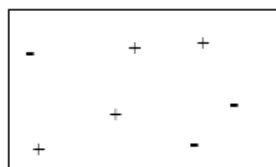
c. The Following diagrams represent training examples for a Boolean-valued concept function plotted in feature space. Show how each of the following algorithms might partition the space based on these examples. (No need for calculations... just give a qualitative answer.)



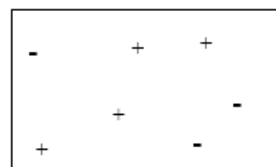
Neural Network with 1 hidden layer of 2 units



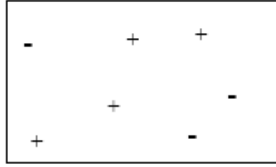
Neural Network with unlimited hidden layers/units



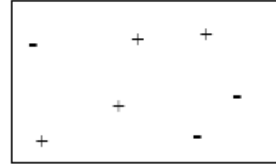
1-Nearest Neighbor



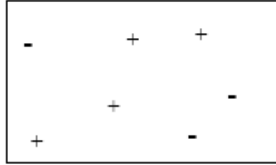
K-means (k=2)



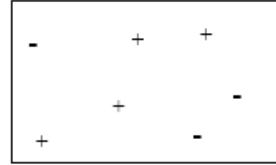
GMM with 2 Gaussian mixtures Models



GMM with 3 Gaussian Mixture Models

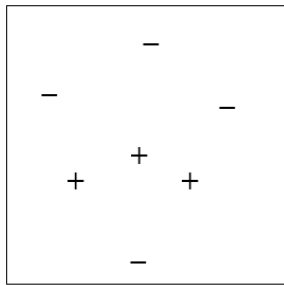


SVM with Linear kernel

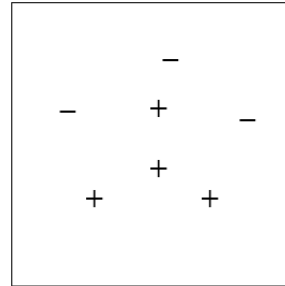


SVM with polynomial kernel

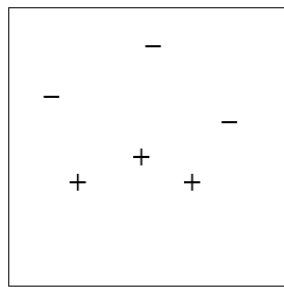
d. The Following diagrams represent training examples for a Boolean-valued concept function plotted in feature space. Consider we have only $(v1.v2)$, $(v1.v2)^2$, RBF $\sigma = 2$, RBF $\sigma = 0.1$ as kernels. For each diagram, write the simplest kernel that correctly classifies all the data, the number of support vectors, and sketch the middle of the separating street.



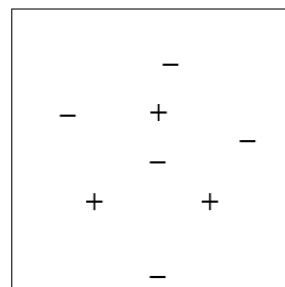
Kernel: #SVs:



Kernel: #SVs:



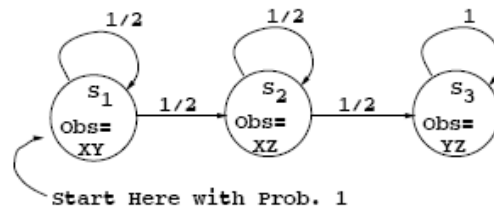
Kernel: #SVs:



Kernel: #SVs:

2. Hidden Markov Models (HMM)

Consider the following HMM:



$a_{11} = 1/2$	$a_{12} = 1/2$	$a_{13} = 0$	$b_1(X) = 1/2$	$b_1(Y) = 1/2$	$b_1(Z) = 0$	$\pi_1 = 1$
$a_{21} = 0$	$a_{22} = 1/2$	$a_{23} = 1/2$	$b_2(X) = 1/2$	$b_2(Y) = 0$	$b_2(Z) = 1/2$	$\pi_2 = 0$
$a_{31} = 0$	$a_{32} = 0$	$a_{33} = 1$	$b_3(X) = 0$	$b_3(Y) = 1/2$	$b_3(Z) = 1/2$	$\pi_3 = 0$

Where

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$$

$$b_i(k) = P(O_t = k | q_t = S_i)$$

Suppose we have observed this sequence: XZXYZZYZZZ

(In long-hand: $O_1 = X, O_2 = Z, O_3 = X, O_4 = Y, O_5 = Y, O_6 = Z, O_7 = Y, O_8 = Z, O_9 = Z$).

Fill in the following table with $\alpha_i(t)$ values, remembering the definition:

$$\alpha_i(t) = P(O_1 \wedge O_2 \wedge \dots \wedge O_t \wedge q_t = S_i)$$

So for example, $\alpha_3(2) = P(O_1 = X \wedge O_2 = Z \wedge O_3 = X \wedge q_3 = S_2)$.

t	$\alpha_1(t)$	$\alpha_2(t)$	$\alpha_3(t)$
1			
2			
3			
4			
5			
6			
7			
8			
9			

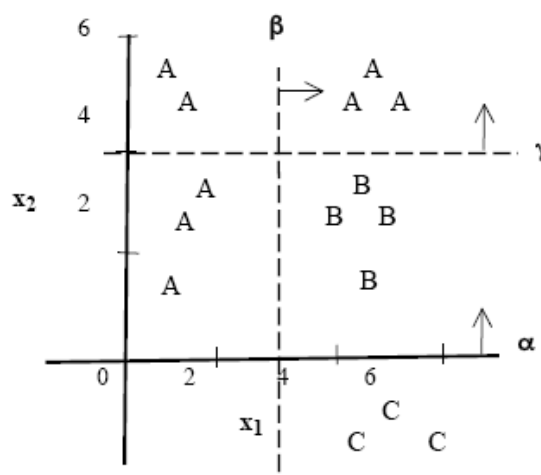
Warning: this is a question that will take a few minutes if you really understand HMMs, but could take hours if you don't!

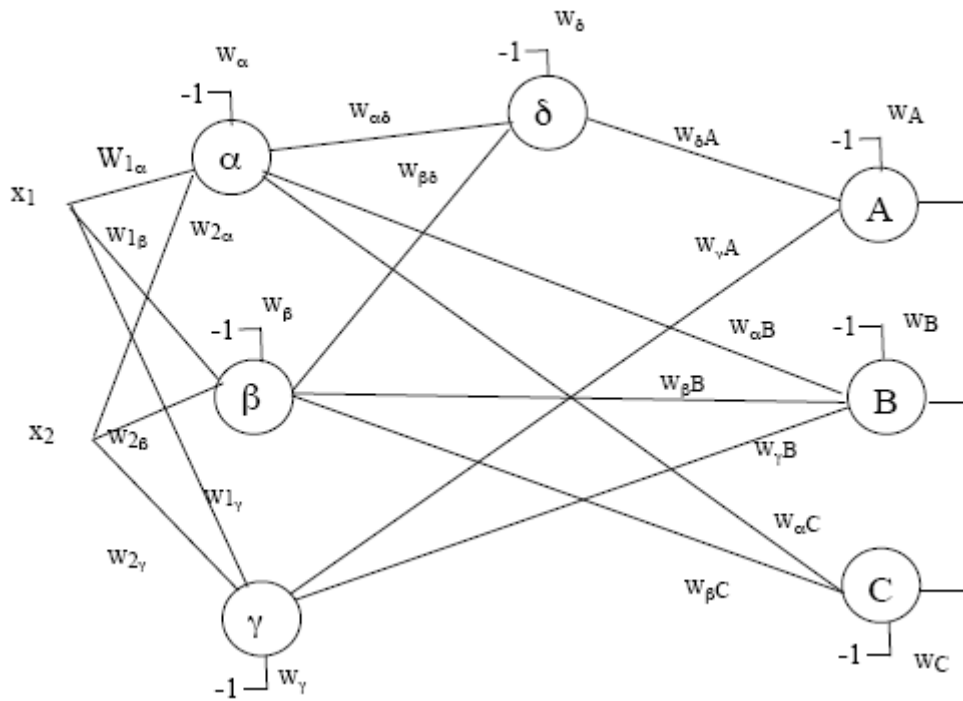
3. Neural Network

a. Consider a neural network with inputs x_1, x_2, \dots, x_n , which are either 0 or 1. This network is specified by giving the weights on the links and the activation function g at the node. Design a network that computes the "majority function" for n input nodes. A majority function should output 1 if at least half the inputs are high, and 0 otherwise.

b. Fill in the missing weights for each of the nodes in the following perceptron network. Make the following assumptions:

- Perceptron output 0 or 1
- A, B, C are classes
- The Lines α, β, γ represent decision boundaries.
- The directions of the arrows shown on the graphs represent the side of each boundary that causes a perceptron to output 1.

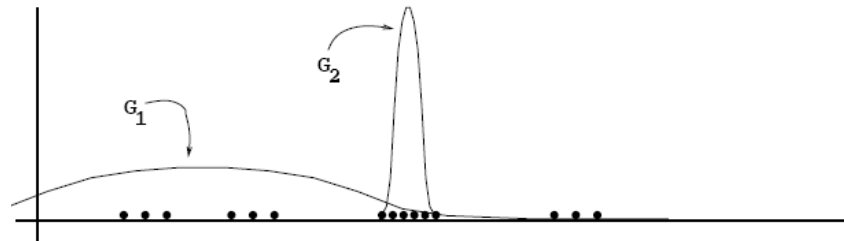




α	β	γ	δ	A	B	C
$w_{1\alpha}$	$w_{1\beta}$	$w_{1\gamma}$	$w_{\alpha\delta}$	$w_{\delta A}$	$w_{\alpha B}$	$w_{\alpha C}$
$w_{2\alpha}$	1	$w_{2\gamma}$	$w_{\beta\delta}$	$w_{\gamma A}$	$w_{\beta B}$	$w_{\beta C}$
w_{α}	w_{β}	w_{γ}	w_{δ}	w_A	$w_{\gamma B}$	w_C

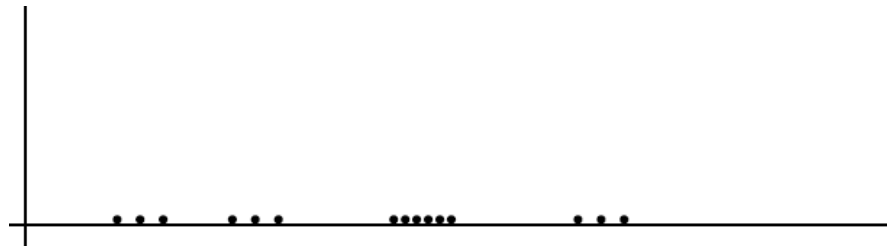
4. Unsupervised learning and EM

The following diagram depicts a mixture of two Gaussians G1 and G2. The mixture model assumption is that each data point is generated by first picking one of the two Gaussians at random (a fifty-fifty chance of either) and then generating the (scalar) value of the data point from the chosen Gaussian.



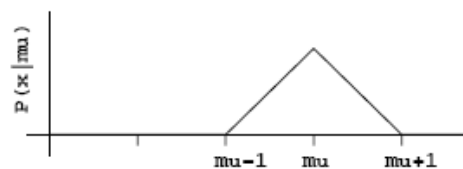
a. For each data point above label which Gaussian, that data point is most likely have been generated by.

b. On the following figure sketch roughly where you believe the means of the two Gaussians would be after one iteration of EM.



A hat distribution is a one dimensional probability distribution with some slightly different properties compared with conventional Gaussian. If x generated from a Hat with mean μ , then its probability density function is given by:

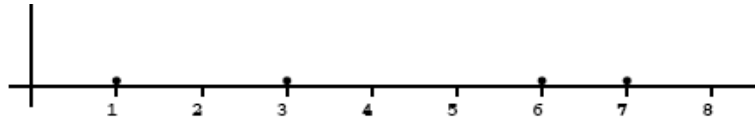
$$p(x | \mu) = \begin{cases} 0 & x \leq \mu - 1 \\ 1 - (\mu - x) & \mu - 1 \leq x \leq \mu \\ 1 - (x - \mu) & \mu \leq x \leq \mu + 1 \\ 0 & \mu + 1 \leq x \end{cases}$$



Assume that the following data was generated by a mixture of three hats:

- Class w_1 is twice as likely as either of the others. It has probability $P(w_1)=1/2$ and mean μ_1 .
- Class w_2 has probability $P(w_2)=1/4$ and mean μ_2 .
- Class w_3 has probability $P(w_3)=1/4$ and mean μ_3 .

Once the class of a data point has been chosen, its value is generated according to that class's hat distribution.



c. Give values for μ_1 , μ_2 and μ_3 that maximize the likelihood of the above data. The values do not have to be integers.

d. What is the numerical value of the likelihood of this data given the values of μ_1 , μ_2 and μ_3 ?