Name and Student ID:

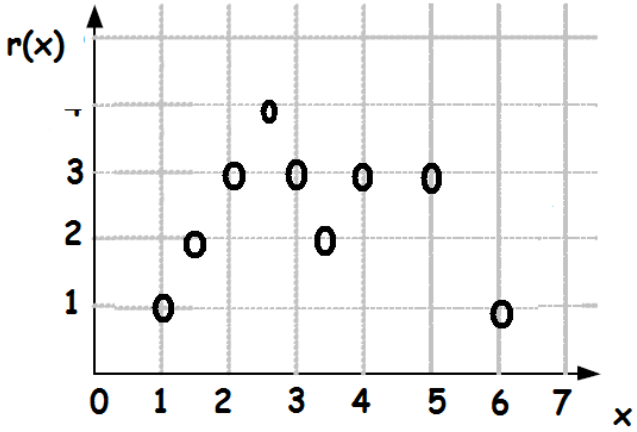
Pattern Recognition and Analysis BBL514E, Jan 17, 2011, Final Exam (25%).

1 30	2 25	3 25	4 20	Total

Name:
Number:
Signature:
Duration: 120 minutes.
Write your answers neatly in the space provided for them. Write your name on each sheet.
Books, notes and cellphones are closed.
Good Luck!

QUESTIONS

QUESTION1) [30 points]



You need to learn an unknown continuous function whose instances are observed above. You are also given the fact that the function is symmetric around x=3.5.
How would you modify the error function and the weight update rule of backprogation algorithm?

When using a neural network, you would have used the following information:

$$y_i = \mathbf{v}_i^T \mathbf{z} = \sum_{h=1}^{H} v_{ih} z_h + v_{i0} \qquad z_h = \text{sigmoid}\left(\mathbf{w}_h^T \mathbf{x}\right) \qquad E(\mathbf{W}, \mathbf{v} \mid \mathcal{X}) = \frac{1}{2} \sum_t \left(r^t - y^t\right)^2$$

$$\Delta v_h = \sum_t \left(r^t - y^t\right) z_h^t$$

$$\begin{aligned} \Delta w_{hj} &= -\eta \frac{\partial E}{\partial w_{hj}} = -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_h^t} \frac{\partial z_h^t}{\partial w_{hj}} \\ &= -\eta \sum_t -\left(r^t - y^t\right) v_h z_h^t \left(1 - z_h^t\right) x_j^t \\ &= \eta \sum_t \left(r^t - y^t\right) v_h z_h^t \left(1 - z_h^t\right) x_j^t \end{aligned}$$

Name and Student ID:

Q1) [30 points]

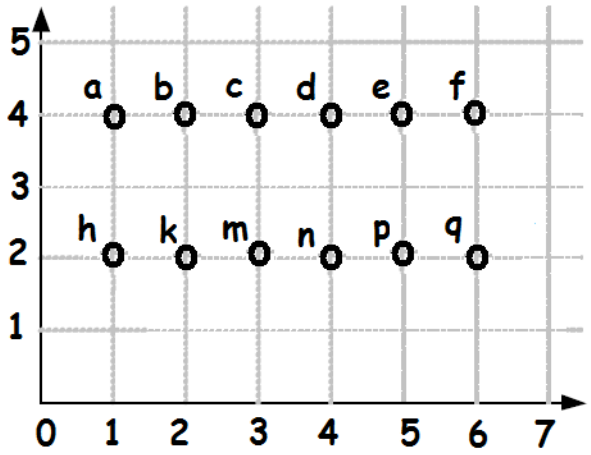
Name and Student ID:

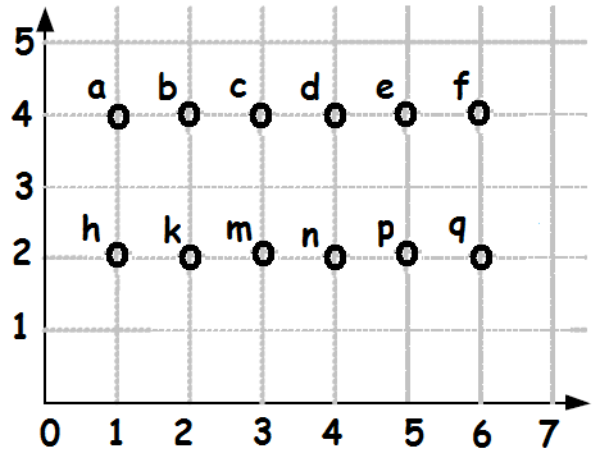
QUESTION 2) [25 points]

Show the resulting clusters when

a) (left graph) Kmeans clustering is used with K=2 and **a** and **f** as the initial points.
b) (right graph) Agglomerative clustering with single linkage is used (i.e. the distance between two clusters is the distance between the two closest data points.
c) Which clustering is more suitable for this dataset?

Use the l_1 norm as the distance measure, i.e. dist(u,v) = $|u_1-v_1|+|u_2-v_2|$





Name and Student ID:

QUESTION 3. [25 points]
Given an HMM $\lambda = (\pi, A, B)$ with state transition probability matrix A, emission probabilities B, initial state probabilities π, and two states and two symbols red and green,

$\pi = [0.2\ 0.8]^T$ $\quad A = \begin{vmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{vmatrix}$ $\quad B = \begin{array}{c|cc|l} & \text{red} & \text{green} & \\ \hline & 0.7 & 0.3 & \text{State1} \\ & 0.4 & 0.6 & \text{State2} \end{array}$

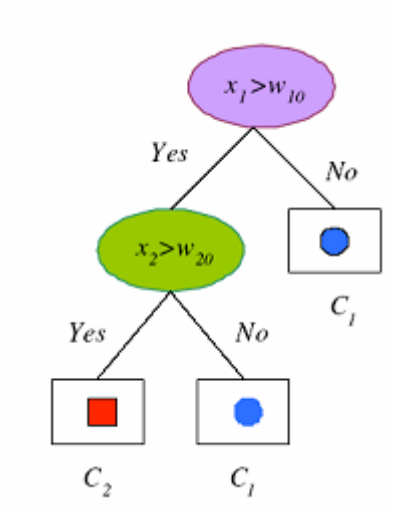
What is the $Pr(O|\lambda)$ where O = {red,red }

Name and Student ID:

Question 4) [20 points] What is (use only the space provided)?:
a) [5 points] The difference between k-means clustering and Gaussian Mixture Model (GMM) clustering?

b) [5 points] k-fold cross validation paired t-test?

c)[10 points] Given the decision tree below and w10=10, w20=20, which class does the data point $[0,0]^T$ belong to?



[8 points] How do you produce a decision tree using using the entropy as the impurity criterion? Assume that you have a classification problem and binary inputs.

Name and Student ID:

Extra sheet