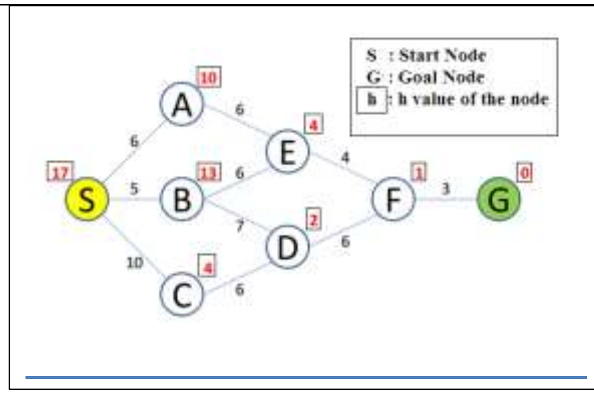
Indian Institute of Information Technology Chittoor

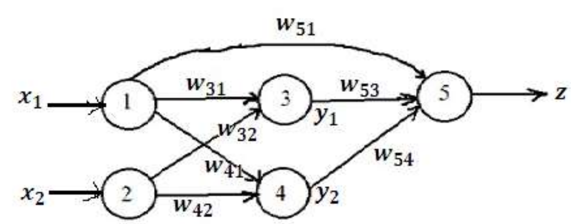
Endsem, **AI**, April 2017; Max marks = 100;

This examination is a closed book and closed notes examination; Calculators are allowed.

CLARITY (NEATNESS) AND PRECISION ARE IMPORTANT (carries 4 marks explicitly). "NOT CLEAR" AND IRRELEVANT ANSWERS MAY ATTRACT NEGATIVE ($-ve$) MARKS. Interleaving (embedding) one answer within some other answer may also attract $-ve$ marks. Each question is for 16 marks.

Any missing information can be appropriately assumed and stated (as part of the answer).

<table>
<tr>
<td>1</td>
<td>(a) Consider the graph given where the number at the top of the node is its h value. Apply A* algorithm (show open and closed lists, for every step, in the form of a table where an entry for each node is represented as $< node,\ parent\_node,\ g + h = f >$) and thus find path from S to G.
(b) Find whether the given heuristic is an admissible one or not.
(c) Find whether the given heuristic is a consistent (i.e., satisfies the monotone property) one or not.</td>
<td>

Note: The given graph is an undirected one. That means, for example, from S one can reach A, from A one can reach S.</td>
</tr>
<tr>
<td>2</td>
<td>For the shown neural network where nodes 1 and 2 are the input neurons and node 5 is the output neuron. It is a conventional feed-forward network except that there is a connection from node 1 to node 5. Let the given input is $(x_1, x_2)^t = (1,1)^t$ and the target output $t = -1$. Let all weights are currently equal to 1. Consider the sigmoid activation function $f(v) = \frac{1}{1+e^{-v}}$, and the learning rate 0.1.
(a) Find the weight updation rule for each of the weights.
(b) Find the updated weights after one iteration.
(c) Pictorially draw the network, both before and after weight updation by showing weights and output.</td>
<td>

{ Use the notation given, which is, weight from node i to node j is w_{ji} . Inputs are x_1, x_2, output of nodes 3, 4 are y_1, y_2. Output is z . Target is t }</td>
</tr>
</table>

3	Consider a two class, two dimensional problem with training set $\{X_1 = (0,0)^t, X_2 = (0,1)^t, X_3 = (1,0)^t, X_4 = (1,1)^t\}$, where X_1, X_2, X_3 belongs to class -1, and X_4belongs to class $+1$. **(a)** Pictorially show the data. **(b)** Formulate the hard linear SVM classification problem{That is, state the optimization problem}. **(c)** List out the K.K.T conditions. **(d)** Solve the K.K.T. conditions with appropriate assumptions (based on the geometry of the problem) and thus find the solution. Show the solution geometrically(pictorially).
4	**(a)** Consider a two class two dimensional problem. The data for each class is from a Gaussian distribution with $\mu_1 = (0,0)^t$, $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, $\mu_2 = (4,4)^t$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Assume $P(\omega_1) = \frac{1}{4}$, and $P(\omega_2) = \frac{3}{4}$. Find the Bayes classifier (i.e., find the formula of the decision surface which divides the feature space in to class regions). Pictorially show the found classifier. **(b)** For a c class problem, prove that the Bayes classifier error, for any underlying distribution of the data, is never worse than the random classifier. Random classifier randomly chooses a class label from the uniform distribution of class labels. That is, each class is chosen with a probability $1/c$.
5	**(a)** Let Σ be the covariance matrix (assume that the data is from a Gaussian distribution). Find x, a unit vector such that the data when projected on to this x, will have the maximum possible variance. This direction is called the first principal component. The Hint is that $x^t \Sigma x$ is the variance along the direction x. But now x must be a unit vector. So the solution has to satisfy $x^t x = 1$. **(b)** Find x, when the covariance matrix $\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$.
6	Consider a one dimensional two class problem where $p(x\|\omega_1) = \begin{cases} 1.5x^2, & -1 \le x \le 1 \\ 0, & Otherwise \end{cases}$, $p(x\|\omega_2) = \begin{cases} 0.75 - 0.75x^2, & -1 \le x \le 1 \\ 0, & Otherwise \end{cases}$. Consider equal priors (i.e., prior probabilities are same). **(a)** One person wants to use a classification rule like $If\ (x \le \theta)\ output\ (\omega_1); else\ output\ (\omega_2)$. Can you express error of this classifier as a function of θ? Show a plot of this function. Can you differentiate this and thus find the optimal value of θ. Is this θ uniquely found? If not, can you give all values of θ? **(b)** Can you state the working of the Bayes classifier in the form of rules? Find the Bayes classifier's error.

-- End --

-- No end for learning --