Roll No:	Name:

INDIAN INSTITUTE OF TECHNOLOGY TIRUPATI

EE5024 – Machine Learning for Image Processing

MARKS: 60 Final Examination TIME: 3 Hours

There are 6 questions and each question carries 10 marks. Answer ALL questions.
Specify clearly the key intermediate steps.

1. Write crisp direct answers to the following sub-questions.

a) An image has 100 lines and 1000 pixels per line. Each pixel can take 512 different values. What are the minimum number of total bits required to store that image.

b) If we quantize the image in part-a) with double resolution (meaning we use twice the number of bits per pixel) and sample it with half the resolution in each dimension, then what is the amount of storage required for storing it?

c) Let the maximum and minimum intensity values of an image are $I_{\max}$ and $I_{\min}$ respectively. Specify a linear intensity transform that maps the image I_1 onto an image I_2 such that the maximum and minimum intensity values of I_2 are L and 0 respectively.

d) What will be the result of repeatedly applying a 3×3 averaging filter to an image a large (infinite) number of times? Assume that the size of the image after convolution is same as the input image by using "replicate" based image padding.

e) Given an image with only 2 pixels and 3 possible values for each pixel. Specify the number of possible different images and the number of possible different histograms.

2. Let $x_1, x_2, \cdots, x_N$ be the independent random samples drawn from an exponential distribution whose pdf is given by:

$$p(x|\lambda) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

a) Compute the expected value of $p(x|\lambda)$ for a given value of the parameter λ, denoted by $E[x]$.

b) Write the expression for the log-likelihood function of λ, denoted by $L(\lambda)$.

c) Compute the maximum likelihood estimate of the parameter λ from $L(\lambda)$, denoted by $\hat{\lambda}$. Also compute $E[x]$ for $\lambda = \hat{\lambda}$.

3. The prior probabilities of samples of two classes denoted by ω_1 and ω_2 are $\frac{1}{3}$ and $\frac{2}{3}$ respectively. The pdf of both classes is given by:

$$P(x|\omega_i) = \begin{cases} 2\theta_i\, x\, e^{-\theta_i x^2}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where θ_i is the parameter associated with class-i.

a) Compute the decision boundary that results in minimum error probability.

b) Let the penalty associated with falsely classifying a sample originating from class-1 to class-2 be 0.5, and from class-2 to class-1 be 1. Compute the decision boundary that minimizes the average risk.

c) If $\theta_1 = 0.005$, and $\theta_2 = 0.5$, compute the corresponding decision boundary values for part-a, and part-b.

4. The training samples of two classes with 2-dimensional feature vectors are as follows: $\omega_1 = \{(0,0)^T,\ (0,-1)^T, (-1,0)^T\}$, and $\omega_2 = \{(1,0)^T,\ (0,1)^T, (1,1)^T\}$. The initial weights for a single perceptron-based linear classification $(w_1, w_2)^T$ are: $(-1,-1)^T$ with an initial bias (w_0) of -0.1. Let $g(\vec{x}) = 0$ be the decision plane. Given that $g(\vec{x}) \geq 0$ for ω_1, and $g(\vec{x}) < 0$ for ω_2.

a) Draw a rough sketch of all the six feature vectors along with the initial decision line. Use different symbols (like 'x' and 'o') for distinguishing features coming from different classes. Also write the equation of $g(\vec{x})$ for the abovementioned initialization.

b) Specify the gradient descent-based equation that can be used for obtaining the updated weight vector $\vec{w}(t+1)$ given the weight vector in the previous iteration ($\vec{w}(t)$), and the constant learning rate parameter (ρ). Give a brief description of its terms.

c) Assume $\rho = 0.5$. Perform one iteration and compute the updated $\vec{w}$ using gradient descent based equation that you mentioned in part-b). Draw the resulting decision plane on the graph that you have already plotted in part-a).

5. The size of the input feature vectors to a multi-layer neural network is 4. It has two hidden layers with 4 neurons in the first layer, and 3 neurons in the second layer. The output layer has one neuron. Recall that the output of a node in the neural network is given by $f(\vec{w}^T\vec{x} + b)$, where $\vec{x}$ is the input to the unit, $\vec{w}$ are the weights on the inputs to that neuron, b is the bias in that unit, and f is the activation function.

a) Draw a block diagram of the neural network described above. Specify the dimensionality of $\vec{w}$ and b for each of the neuron in the first hidden layer, second hidden layer and the output layer of the abovementioned neural network.

b) Write the set of equations that establishes the relationship between the output of this network and the input feature vector in the testing data.

c) Specify the equation $f(x)$ for sigmoid activation function. Compute the derivative of $f(x)$, and express it in terms of $f(x)$.

6.

a) Assume that we need to minimize $||\vec{w}||^2$ subjected to the constraints that $(\vec{w}^T\,\vec{x_j} + w_0) \geq 1$ for all integer values of j from 1 to N, where $\vec{w}$ and w_0 are the varying parameters. Instead of solving this primal problem, use Lagrange multipliers and formulate an equivalent dual problem.

b) Find the points on the circle $x^2 + y^2 = 80$ which are the closest to and the farthest from the point $(1,\ 2)$ using Lagrange multipliers.