# Data Mining and Machine Learning

## Final Examination, II Semester, 2024–2025

Date : 30 April, 2025
Duration : 3 hours

Marks : 40
Weightage : 40%

1. Explain how to cluster points using a mixture of Gaussians. Can this also be used to detect outliers?

*(5 marks)*

2. A die has six faces labelled $\{1, 2, 3, 4, 5, 6\}$. The die is fair if each value $j \in \{1, 2, \ldots, 6\}$ appears with equal probability $\frac{1}{6}$ when the die is rolled. Otherwise, it is said to be unfair.

   You have two dice, $D_1$ and $D_2$, which may be unfair. You are given a sequence of 10,000 values from $\{1, 2, \ldots, 6\}$ where each outcome corresponds to choosing one of $D_1$ and $D_2$ uniformly at random and then rolling it.

   For $i \in \{1, 2\}$ and $j \in \{1, 2, \ldots, 6\}$, let $p_{ij}$ denote the true probability that dice $D_i$ displays the value $j$ when rolled.

   Describe, in algorithmic pseudocode, an iterative procedure to estimate $\{p_{11}, p_{12}, \ldots, p_{16}\}$ and $\{p_{21}, p_{22}, \ldots, p_{26}\}$.

*(5 marks)*

3. The 0–1 loss function assigns a cost of 1 to every misclassified input and a cost of 0 to every correctly classified input. This loss function is minimized when the model makes no errors on the training data. Explain with respect to linear separators why the 0–1 loss function is not always adequate to learn a good model.

*(5 marks)*

4. Explain, using a suitable example, why relaxing the penalty for misclassification can result in a wider margin for a soft margin SVM.

*(5 marks)*

5. The radial basis kernel is given by $K(x, z) = e^{-\gamma|x-z|^2}$. Explain how this kernel is related to the idea of using similarity features for computing a linearly separable representation of a given dataset.

*(5 marks)*

6. A dataset has been created with images of handwritten uppercase letters $\{A, B, \ldots, Z\}$. Each input is a $30 \times 30$ pixel grayscale image and there are 500 training inputs for each of the 26 possible input letters.

   A dense, acyclic, layered neural network is being trained to classify inputs from this dataset. There are two hidden layers. Each hidden layer has 100 neurons with a RELU activation function. The output layer computes a softmax across the 26 letters.

   How many parameters need to be trained in this network using backpropagation? Explain your answer.

*(5 marks)*

*... (More questions on page 2)*

7. In a nuclear power station, an alarm is triggered when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core of the reactor. Consider the boolean variables $A$ (alarm sounds), $F_A$ (alarm is faulty), and $F_G$ (gauge is faulty) along with multivalued variables $G$ (gauge reading) and $T$ (actual core temperature) that take values from a set $V = \{t_1, t_2, \ldots, t_k\}$.

(a) Draw a Bayesian network for this scenario, given that the gauge is more likely to fail when the core temperature gets too high. Explain the structure of your network.

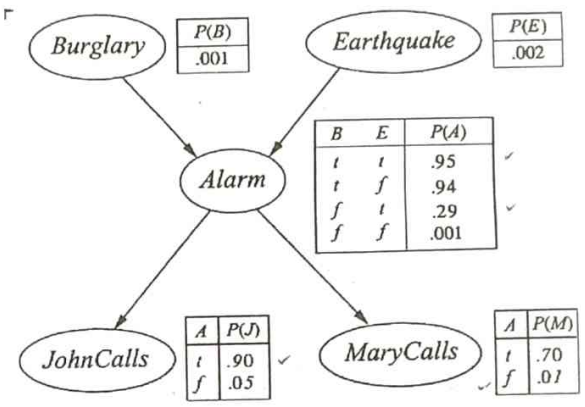
(b) Suppose $G$ and $T$ each take just two values, normal and high. Assume that the gauge gives the correct temperature with probability $x$ when it is working and with probability $y$ when it is faulty. Describe the conditional probability table for $G$.

*(5 marks)*

8. For the network on the right, we want to use likelihood weighted sampling to estimate the probability that there has been an earthquake, given that John has called but Mary has not called.

(a) List out all the types of samples that will be generated by likelihood weighted sampling.

(b) Compute the likelihood weight associated with each type of sample. (You can express the weights as arithmetic expressions, without calculating the final value.)




*(5 marks)*