Data Mining and Machine Learning

Final Examination, II Semester, 2023–2024

Date	3 May, 2024	Marks	40
Duration	3 hours	Weightage	40%

1. There are three biased coins c_1, c_2, and c_3. You are given a sequence of 1000 coin tosses, where each outcome corresponds to tossing one of $\{c_1, c_2, c_3\}$, chosen uniformly at random. Let $\{p_1, p_2, p_3\}$ be the probabilities of heads for the coins $\{c_1, c_2, c_3\}$, respectively. You have prior information that p_1 is less than 0.5 and p_2 and p_3 are greater than 0.5. Describe, in algorithmic pseudocode, an iterative procedure to estimate $\{p_1, p_2, p_3\}$. *(5 marks)*

2. Explain how to cluster points using a mixture of Gaussians. Can this also be used to detect outliers? *(5 marks)*

3. Explain how clustering can be used for image segmentation — that is, to identify objects in an image. *(5 marks)*

4. 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 the perceptron algorithm why the 0–1 loss function is not always adequate to learn a good model. *(5 marks)*

5. (a) For $z = wx + b$, how does the shape of the sigmoid function $\sigma(z) = (1 + e^{-z})^{-1}$ vary with w and b?

   (b) Given two input features x_1, x_2, explain how to construct a neural network to approximate a "rectangular box" function $g(x_1, x_2)$ with height h for $\ell_1 \leq x_1 \leq r_1$ and $\ell_2 \leq x_2 \leq r_2$. In other words, the function to be approximated is the following.

$$g(x_1, x_2) = \begin{cases} h & \text{if } \ell_i \leq x_i \leq r_i, i \in \{1, 2\} \\ 0 & \text{otherwise} \end{cases}$$

(5 marks)

6. Consider a neural network that is layered and completely connected. Suppose we initialize two nodes n_1 and n_2 from the same layer with the same biases and same weights on incoming and outgoing edges. What can you say about the final weights and biases that will be learned for n_1 and n_2 through backpropagation? What can you conclude about initialization strategies for such networks? *(5 marks)*

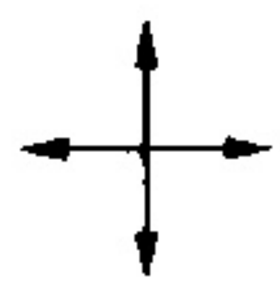
7. Two astronomers independently count stars in the same region of the sky using their telescopes. The region has N stars. The counts reported by the astronomers are M_1 and M_2, respectively. Each astronomer has a small probability of miscounting the stars by ± 1. It is also possible that their telescopes are faulty and do not focus properly, denoted by boolean events F_1 and F_2, respectively. With a faulty telescope, an astronomer may undercount by as many as 3 stars.

   (a) Draw a Bayesian network to represent the relationship between N, M_1, M_2, F_1 and F_2.

   (b) Suppose $M_1 = 12$ and $M_2 = 11$. What are the possible values of N for each of the different combinations of F_1 and F_2?

(5 marks)

8. Consider the 4×4 grid-world to the right. The non terminal states are $\{1, 2, \ldots, 14\}$ and the terminal states are the shaded squares. There are four actions, $\{\text{up}, \text{down}, \text{left}, \text{right}\}$, which result in a deterministic move in the given direction. A move that would take the agent off the grid leaves the position unchanged. The reward is -2 for any transition that results in a change of position. A move off the grid that does not change the position has a reward of -1. Formally, $r(s, a, s') = -2$ if $s \neq s'$ and $r(s, a, s') = -1$ if $s = s'$.



actions

	1	2	3
4	5	6	7
8	9	10	11
12	13	14	

(a) Consider the uniformly random policy π that chooses each of the four directions with equal probability. Assume we start with an initial value $v(s) = 0$ for each state s. Compute one iteration of v_π.

(b) Describe the new policy after applying policy improvement based on this one step computation of v_π.

(5 marks)