

Time: 2.5 hours

2023-24 Semester I

Max mark: 35

Please read the following instructions carefully.

- There are 8 questions printed on **2** pages.
- **No marks** for providing just expressions/answers unless accompanied with correct justification and/or derivation.
- In case of any doubt, make your assumption, write it clearly and continue.

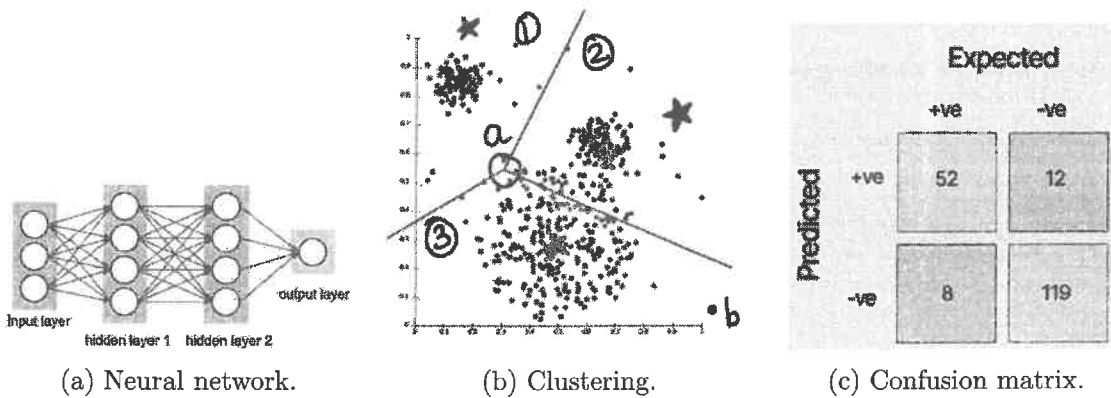


Figure 1: All the figures referenced in the question paper are here.

1. Consider the neural network shown in Figure 1a. (CO2, CO3)

- (a) (4 points) How will you represent the function learnt by this neural network using an equation? To help you out, the computations of the last layer are provided below. You are supposed to replace the proper variables in this equation to finally make it look like a function of the learnable parameters, the input (x) and activation functions  $g^{[1]}$  and  $g^{[2]}$  only. Your final answer has to be one equation only.

$$\hat{y} = \sigma(W^{[3]}a^{[2]} + b^{[3]})$$

- (b) (2 points) After training for long enough time on the neural network as provided in Figure 1a, you realize that the training accuracy is 78% and validation accuracy is 72%. However, you know that one of your friends has already obtained a training accuracy of 98% and validation accuracy of 96% on the same task and the same dataset. What changes will you make in the provided network to get closer to the accuracy attained by your friend?
- (c) (2 points) Consider another scenario, in which you have reached a training accuracy of 99% and validation accuracy of 80% on the network in Figure 1a. What changes will you make to the neural network to improve the validation accuracy?

2. (2 points) Why is mean squared error not used as a cost function for logistic regression (crossentropy is used instead)? (CO3)
3. (2 points) You are provided a dataset of weight of 1 billion aliens on the exoplanet K2-18b. This is the first column of your csv file. The second column of this csv file is a label stating whether the alien is obese (overweight) or not. You also know that in this dataset all the aliens with weight greater than 80kg are labeled as obese and all the aliens with weight less than or equal to 80kg are labeled as not obese. There is not even a single exception to this rule in the dataset. Argue which machine learning algorithm you will use for this problem, if at all. Clearly state your argument. (CO1,CO2)
4. Consider a dataset after running k-means clustering in Figure 1b. There are three clusters numbered as 1, 2 and 3. You are also provided three points marked by stars. Point *b* is not the part of the original dataset on which the shown clustering was done. (CO3, CO5)
  - (a) (2 points) How will you interpret the intersection of the three lines at point *a*?
  - (b) (2 points) Do all the three stars represent the centroid of their respective clusters? Comment on the possibility of each.
  - (c) (2 points) If a new point *b* is provided at the test time, which cluster will it go to?
  - (d) (2 points) Assume that the dataset has one point more than the points in the provided figure. Let that extra point be *p*. The x-coordinate of the point *p* is the sum of the x-coordinates of all the points in cluster 3. Assume that the y-coordinate of point *p* is 0, what would be the impact of this point if we were to re-run k-means from the start?
5. (3+2 points) Assume you are provided a one dimensional data. On looking at the plot of the dataset, it appears there are three clusters. What are the parameters to be learnt in a Gaussian mixture model? How is it different from k-means clustering? (CO2, CO5)
6. You are provided a 500 dimensional data in the form of a csv file, i.e., the csv file has 500 columns. Each row represents an example in the dataset. You want to convert it to a 2 dimensional data. You know that PCA can be used to achieve the goal. State whether the following are true or false, with justification. (CO2, CO3, CO5)
  - (a) (2 points) PCA will choose any 2 columns and will discard the remaining 498 columns.
  - (b) (2 points) There will be no information loss while converting a 500 dimensional feature vector to two dimensional feature vector.
7. (2 points) Write gradient descent algorithm. (CO1, CO2, CO3)
8. (4\*1 points) You are provided a confusion matrix in Figure 1c. Calculate accuracy, recall, precision and f1-score from it. (CO3, CO5)