B.Sc. Engg. CSE 6th Semester **05 April 2022**

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION **SUMMER SEMESTER, 2020-2021**

DURATION: 3 HOURS **FULL MARKS: 150**

CSE 4621: Machine Learning

Programmable calculators are not allowed. Do not write anything on the question paper.

Answer **all 6 (six)** questions. Marks of each question and corresponding CO and PO are written in the right margin with brackets. Symbols have their usual meaning.

---

1. a) Suppose the following 2D data is given. 20 [CO1, PO1]

Table 1

x_1	x_2
3	3
4	4
5	5

Show all numeric calculations required to perform Principal Component Analysis (PCA) to transform the original data into 1D data with maximum variance.

Note: One of the Eigen values of the covariance matrix is 0, and corresponding Eigen vector is $[-0.7071\ 0.7071]^T$.

b) Is there any difference if we choose the number of principal components using either 'the maximum projection error allowed' or 'minimum variance retained'? Justify your answer. 5 [CO2, PO2]

2. a) Identify the strengths and weakness of elbow method in determining the number of clusters in a dataset? 5 [CO2, PO2]

b) Does k-medoids clustering technique produce convex-shaped clusters? Explain your answer. 7 [CO1, PO1]

c) Compare between k-means and k-medoids clustering techniques. 8 [CO2, PO2]

d) Design a criterion function J for partitioning, where for each class a distance measure can be used between samples p, instead of using cluster center c_i. 5 [CO3, PO3]

3. a) Consider a Support Vector Machine and the following training data for a two-class problem given in Table 2:

Table 2

class	x_1	x_2
+	1	1
+	2	2
+	2	0
−	0	0
−	1	0
−	0	1

i. After plotting these six training points (use graph paper), construct the weight vector for the optimal hyperplane, and the optimal margin width. 10+5 [CO1, PO1]

ii. If you remove one of the support vectors does the size of the optimal margin decrease, stay the same, or increase?

[Note: You do not need to calculate the solutions by solving, rather find the answers from inspecting the graph.]

b) Suppose the test data in a two-class problem is not linearly separable and noisy with outliers. Which concepts can you employ to make the SVM classifier work with better generalization? Explain in brief with changes in the objective function. 10 [CO3, PO3]

4. a) Compare between Generative and Discriminative models. 5 [CO2, PO2]

b) Suppose you are training a robot in a lumber yard, and the robot must learn to discriminate Oak wood from Pine wood. You choose to learn a Naïve Bayes classifier with the following data in Table 3: 4+1 [CO1, PO1]

Table 3

Density	Grain	Hardness	Class
Heavy	Small	Hard	Oak
Heavy	Large	Hard	Oak
Heavy	Small	Hard	Oak
Light	Large	Soft	Oak
Light	Large	Hard	Pine
Heavy	Small	Soft	Pine
Heavy	Large	Soft	Pine
Heavy	Small	Soft	Pine

Consider a new sample (Density=**Light**, Grain=**Small**, Hardness=**Soft**)^T. Calculate the posterior probability for each class and classify the sample

c) For the Naïve Bayes classifier, the decision rule *f(x)* can be written as follow, where the sample will be classified to the positive class (i.e., $y=1$) if $f(x)>0$: 5 [CO3, PO3]

$$f(x) = \log \frac{P(y=1 \mid x)}{P(y=0 \mid x)}$$

Can the decision rule be formulated similarly for multiclass Naive Bayes? Explain why.

d) During decision tree generation for classification, instead of taking a binary split for the numeric attribute, can we use ternary split using two thresholds w_{ma} and w_{mb}? In other words, three potential branches where samples can take j-th branch according to the following conditions: 7+3 [CO3, PO3]

$$x_j < w_{ma} ; w_{ma} \le x_j \le w_{mb} ; x_j > w_{mb}$$

Propose a modification of the tree induction method along with impurity measure to learn those two thresholds. What are the advantages of performing ternary split over binary?

5. a) What is the motivation behind 1×1 convolution? How can it help in reducing the computation cost? Give an example scenario 5 [CO1, PO1]

b) Consider the following LeNet-5 model in Figure 1. If we replace the average-pool with max-pool layers, determine the changes you might see? 7 [CO2, PO2]

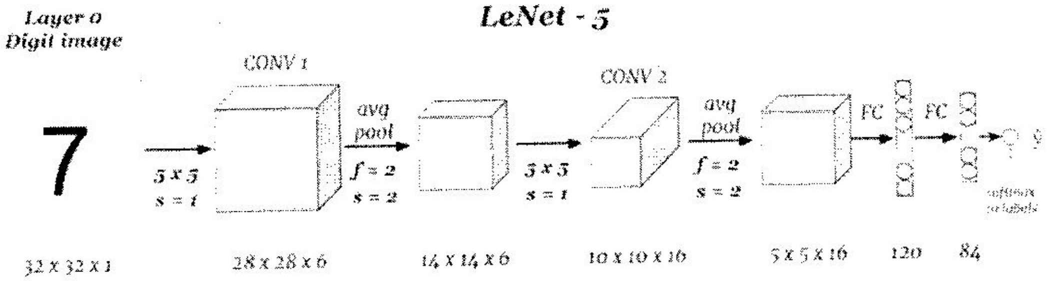

Figure 1.

c) Examine the benefits of using skip connections in Convolution Neural Network (CNN). How do we incorporate such skip connections and in which cases? 6 [CO2, PO2]

d) You come up with a CNN classifier as shown in Figure 2. For each layer, measure the number of weights, number of biases and the size of the associated feature maps. 7 [CO1, PO1]

The notation follows the convention:

- CONV-K-N denotes a convolutional layer with N filters, each them of size $K \times K$, Padding and stride parameters are always 0 and 1 respectively.
- POOL-K indicates a $K \times K$ pooling layer with stride K and padding 0.
- FC-N stands for a fully-connected layer with N neurons.

Layer	Activation map dimensions	Number of weights	Number of biases
INPUT	128 × 128 × 3	0	0
CONV-9-32			
POOL-2			
CONV-5-64			
POOL-2			
CONV-5-64			
POOL-2			
FC-3			

Figure 2.

6. a) For large batch sizes, the number of iterations does not change much as the batch size is increased. Explain this statement. 5 [CO1, PO1]

b) Why do we need a Regularization term in the cost function? Compare between L1 and L2 Regularization. 2+5 [CO2, PO2]

c) The standard form of L2-regularized loss function for linear regression is:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right)^2 + \frac{\lambda}{m}\theta^T\theta$$

i. Suppose you have accidentally defined: $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^i) - y^i\right)^2 + \frac{\lambda}{m}Y^T Y$. What kind of regulization effect will you have?

ii. Suppose we use the correct expression but accidentally choose λ<0. Will you either have overfitting or underfitting? Justify your answer. 4+4 [CO2, PO2]

d) If the following first-order condition is true:

$\forall x, y \in \operatorname{dom} f,\; f(y) \geq f(x) + \left[\nabla f(x)\right]^T.(y - x)$, then determine that the function f is convex. 5 [CO3, PO3]