Friedrich-Alexander-Universität Erlangen-Nürnberg
Faculty of Engineering
Chair of Multimedia Communications and Signal Processing
Prof. Dr. Vasileios Belagiannis

Machine Learning for Signal Processing
Test (Not-graded), Winter Semester 2023/24

Duration: 40 Minutes Date: 30.01.2024

Name: _____

Matriculation Number : _____

Study Program: _____

Degree (Bachelor/Master): _____

Information

- Only the provided exam shall be used for completing the tasks. If you need more blank sheets, please contact the examiner.
- Please include your matriculation number on each sheet.
- Please reply to each question in a separate answer box. Boxes are provided below the questions.
- Please write cleanly and legibly. Sections that we cannot read will be scored zero.
- Results without any comprehensible justification or without calculations for arriving at the answers can not be graded.
- Allowed items: only a permanent pen (not of red colour).
- The total number of pages is 7.

Point Distribution

Task	1	2	3	Total
Points:	14	15	16	45
Obtained Points:				

Grade: _____

Task 1 Multi-Topic Questions

1. Reply the following questions using the empty boxes.

   (a) (2 points) What is the difference of the lasso regression from the standard linear regression?

   (b) (3 points) Consider the target value y and the linear model $f(x)$ with x as input. Given m training samples, define the mean squared error loss function.

   (c) (1 point) What do you call the approach of ensemble learning, where a number of weak classifiers are combined to produce a powerful model?

   (d) (3 points) What is a major disadvantage of the Empirical Risk Minimization? (A) List two gradient-based optimizers (B).

   (e) (2 points) What are the two major limitations of the Linear Discriminant Anal-

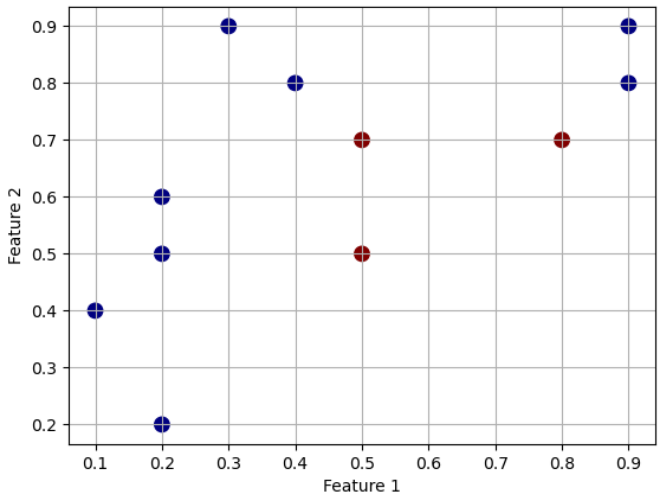
ysis?
Reply to each question in a separate answer box.

(f) (3 points) Consider a convolutional network trained for a classification task on the ImageNet dataset. The input to the network is RGB images of size 224×224. How many parameters are present in the first convolution layer of the network that uses 16 filters of size 5×5 with a stride of 1, and each filter has a bias term? Assuming no padding is used, what will be the dimensions of the convolution layer output? Express the result in the form: feature maps $\times$ height $\times$ width.

Task 2 Decision Trees

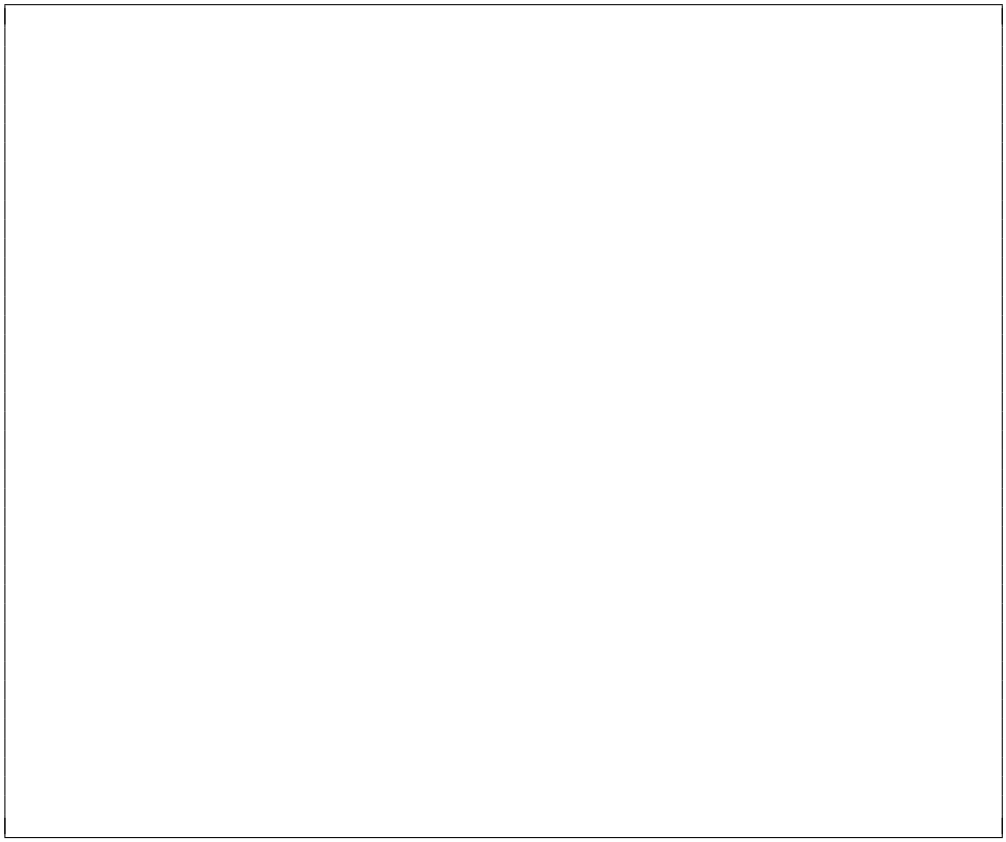
2. Reply the following questions using the empty boxes.

   (a) (5 points) In decision trees, evaluating the possible split is based on the purity of a node. One purity measure is the Gini score that is defined as: $g(\mathbf{p}) = \sum_{i=1}^{k} p_i(1-p_i)$ with $\mathbf{p} = \{p_1, \ldots, p_k\}$ and p_i the fraction of samples that correspond to class i of total k classes. The score is 0 if all samples are from the same class and it increases as the class mix becomes uniform. Calculate the Gini score for the samples drawn in the figure. There are 11 samples and 2 classes in total.

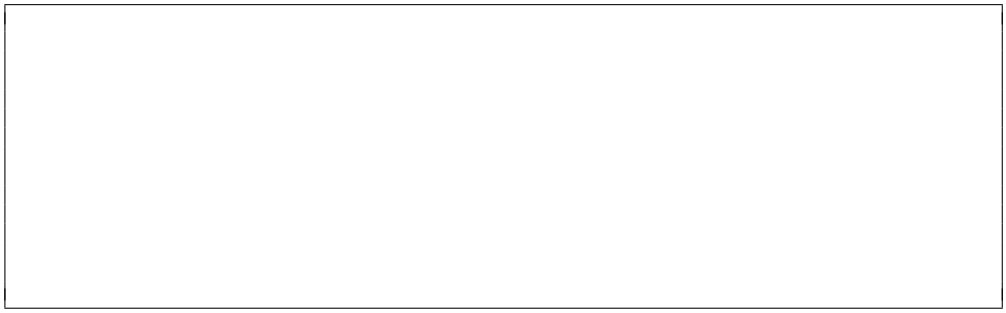


   (b) (5 points) The decision tree formation is a recursive process that splits the training samples into subsets. The criterion to split the node data is to minimise the Gini impurity. The goodness-of-split can be defined as: $F(s,n) = g(\mathbf{p}) - S_l * g(\mathbf{p}_l) - S_r * g(\mathbf{p}_r)$ where s is the split function for node n. S_l corresponds to the fraction of samples at the left node and S_r to the right one accordingly. The higher the value of $F(s,n)$, the better the split is. Grow a decision tree until

you reach leaves (end nodes) with samples of the same class. The goal is to seek for the best split function per node. Report for each node the goodness-of-split and the selected split function.



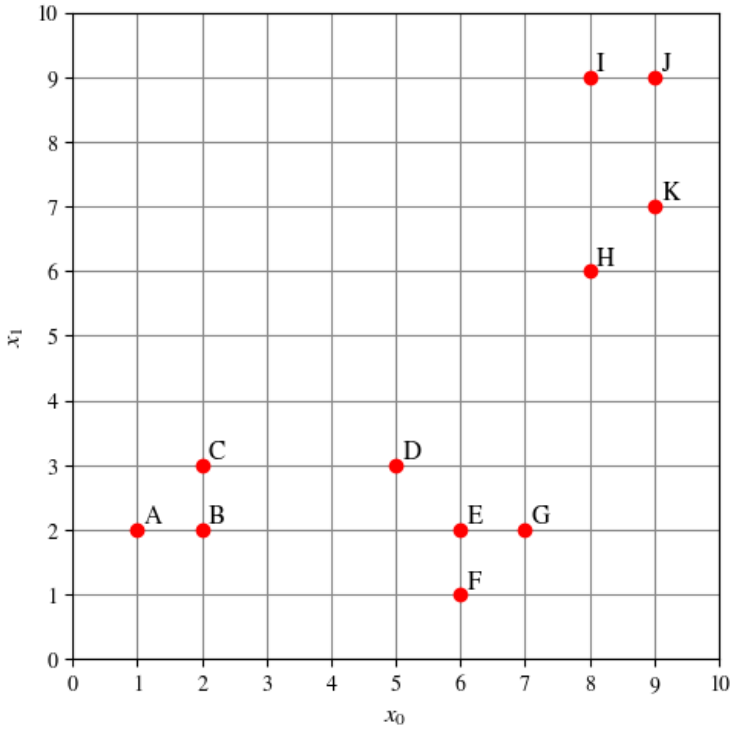
(c) (5 points) Based on the constructed tree classify the following samples: (0.3, 0.2), (0.6,0.5), (0.3, 0.8), (0.6,0.7).



Task 3 K-means clustering

3. Reply the following questions using the empty boxes.

   The scatterplot of a dataset containing 11 samples is given in the figure below. The samples are annotated alphabetically. The goal of this task is to use the K-means clustering algorithm with Euclidean distance as the distance metric to divide the dataset into 3 clusters. Assume that the points E, I and H were randomly selected as cluster centers in the initialization step. Perform the first iteration of the K-Means algorithm and answer the following questions.



(a) (8 points) List the labels for the points that belong to each cluster with cluster centers from initialization step.

(b) (5 points) What are the new cluster centers after the first iteration of the algorithm?

(c) (3 points) List the labels for the points that belong to each cluster with cluster centers after the first iteration.