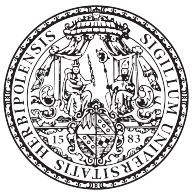


Introduction to Informatics
WiSe 2024/2025

Prof. Dr. Ingo Scholtes
Chair of Informatics XV
University of Würzburg

Exercise Sheet 10

Solutions

1. **Multiple Choice**
   (a) What is the maximum number of steps required by the binary search algorithm to locate an element in a sorted array of length 16?
      A. 8
      B. 4
      C. 16
   (b) Which of the following statements about the Control Unit (CU) in a van Neumann architecture are correct
      A. The CU is responsible for performing arithmetic and logical operations.
      B. The Control Unit updates the Program Counter (PC) to point to the next instruction.
      C. The Control Unit loads the next instruction from memory into the Instruction Register
   (c) Which of the following IP addresses are valid?
      A. 192.168.0.256
      B. 2001:0db8:85a3:0000:0000:8a2e:0370:7334
      C. 10.0.0.1
   (d) What statements about an algorithm that has a time complexity of n^3 are true?
      A. It is a polynomial algorithm.
      B. If the input size doubles, the runtime will increase by a factor 8
      C. It is faster than an exponential algorithm with complexity 2^n

2. **Number Systems**
   (a) Convert the given decimal numbers into their binary and hexadecimal equivalents.
      (i) 34
      (ii) 255

   Solution:

   (i) $34 = 100 = 0x22$

   (ii) $255 = 2^8 + 2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = 11111111 = 0xFF$

   (b) Convert the given binary numbers into their decimal and hexadecimal equivalents.
      (i) 1101
      (ii) 10010010

   Solution:

   (i) $1101 = 13 = 0xD$

   (ii) $10010010 = 146 = 0x92$

(c) Convert the following hexadicial numbers into their binary and decimal equivalents.

(i) 0x2A

(ii) 0x1E

Solution:

(i) 0x2A = 0010 1010 = 42

(ii) 0x1E = 0001 1110 = 30

3. **Logic Operators**

(a) Give the truthtable for the formula

A **AND** (B **OR** ¬C)

Solution:

A	B	C	(B **OR** ¬C)	A **AND** (B **OR** ¬C)
0	0	0	1	0
0	0	1	0	0
0	1	0	1	0
0	1	1	1	0
1	0	0	1	1
1	0	1	0	0
1	1	0	1	1
1	1	1	1	1

(b) Give a formula for the Boolean function represented by the following truth table.

A	B	C	???
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

Solution: $(A$ **AND** $B)$ **OR** $\neg C$

(c) Show that the logical operation **XOR** (with two inputs A and B) can be constructed as Boolean function that only uses **AND, OR and NOT** operations.

Solution: A **XOR** $B = (A$ **AND** $\neg B)$ **OR** $(\neg A$ **AND** $B)$

4. **Ciphers**

(a) The following cipher is encrypted with the Caesar cipher with shift key 3. Decrypt the cipher: `hohskdqw`

Solution: elephant

(b) Use the Vigenère cipher to encrypt the following message with the key `'ABC'`:
`Hello World!`

Solution: `hfnlp yosnd`

5. **Sorting Algorithms**

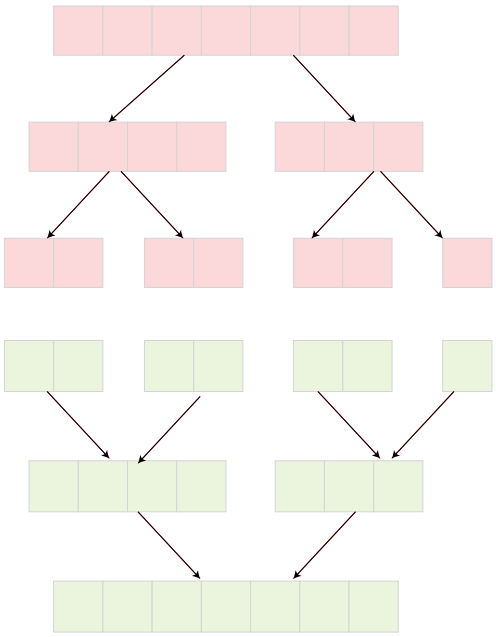
(a) Use the following algorithm on the list `[54,38,2,49,97]` and write down the result of each step in the for loop.

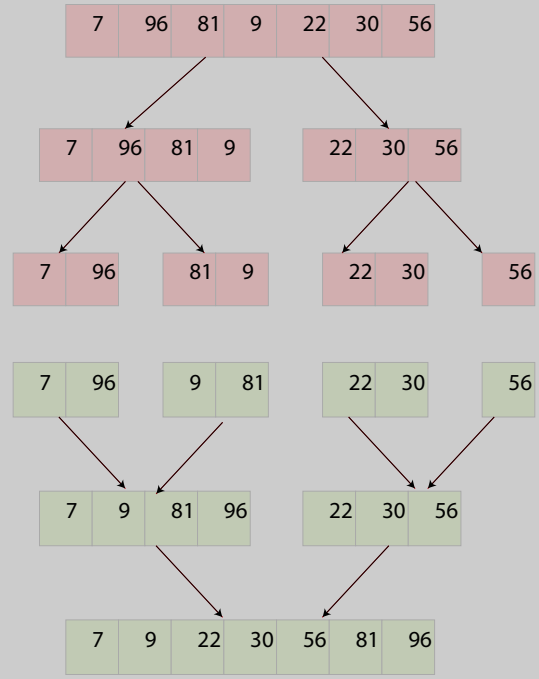
```
def function(arr):
    n = len(arr)
    for i in range(n):
        for j in range(0, n-i-1):
            if arr[j] > arr[j+1]:
                arr[j], arr[j+1] = arr[j+1], arr[j]
    return arr
```

Solution: This is the bubble sort algotithm. The seps are as follows:

- `[54,38,2,49,97]`
- `[38,54,2,49,97]`
- `[38,2,54,49,97]`
- `[38,2,49,54,97]`
- `[2,38,49,54,97]`

(b) Perform a merge sort for the unsorted list `[7, 96, 81, 9, 22, 30, 56]` using the given diagram.





Solution:

6. **Classification Problems**

   (a) You are given the ground truth labels and predicted labels for a binary classification problem.

- Ground truth labels: `[1, 0, 1, 1, 0, 1, 0, 0, 1, 0]`
- predicted labels: `[0, 0, 0, 1, 0, 1, 0, 0, 1, 0]`

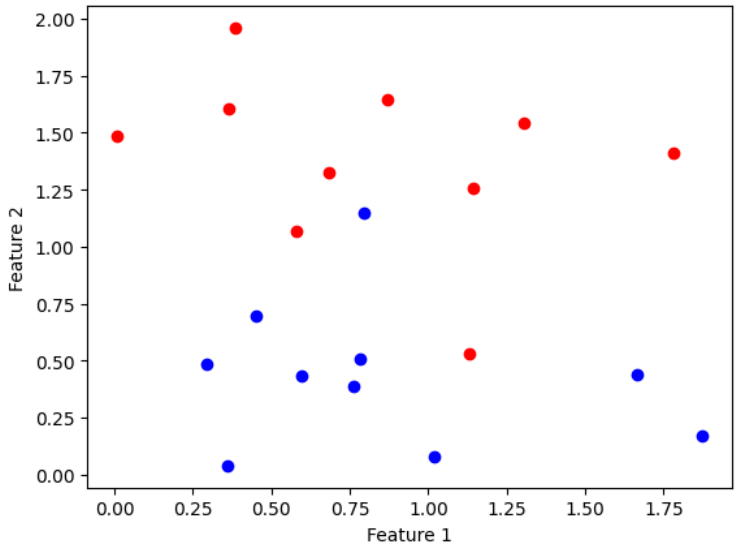
Calculate the confusion matrix, accuracy, precision, recall, and F1 score.

Solution:

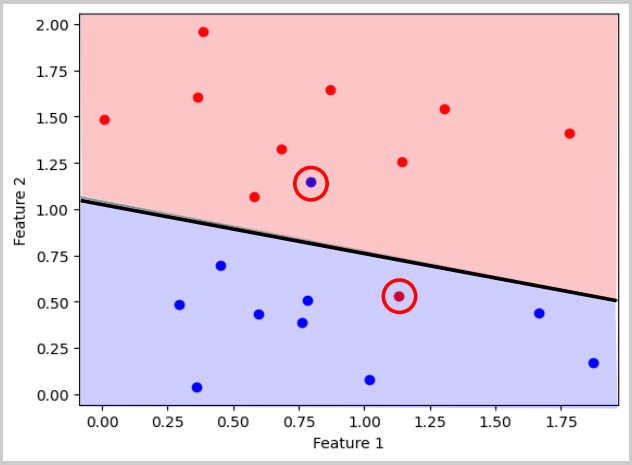
		Ground Truth	
		Positive	Negative
Predicted	Positive	3	0
	Negative	2	5
		5	5

- Accuracy $= \frac{TP+TN}{P+N} = \frac{3+5}{5+5} = \frac{8}{10} = 80\%$
- Precision $= \frac{TP}{TP+FP} = \frac{3}{3+0} = \frac{3}{3} = 100\%$
- Recall $= \frac{TP}{P} = \frac{3}{5} = 60\%$
- F1 Score $= \frac{2TP}{2TP+FP+FN} = \frac{2*3}{2*3+0+2} = \frac{6}{8} = 75\%$

(b) Given the following 2-dimensional dataset with two classes, what is the maximal accuracy that a logistic regression model can achieve?



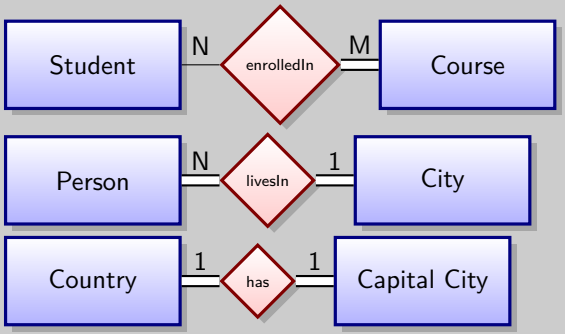
Solution: Logistic regression always produces a linear decision boundary. In this case, it is impossible to draw a linear boundary that perfectly separates the two classes, meaning that at least one data point from each class will always be misclassified. As a result, the maximum achievable accuracy is 90%, with 18 out of the 20 data points correctly classified.



7. **ER Diagrams**

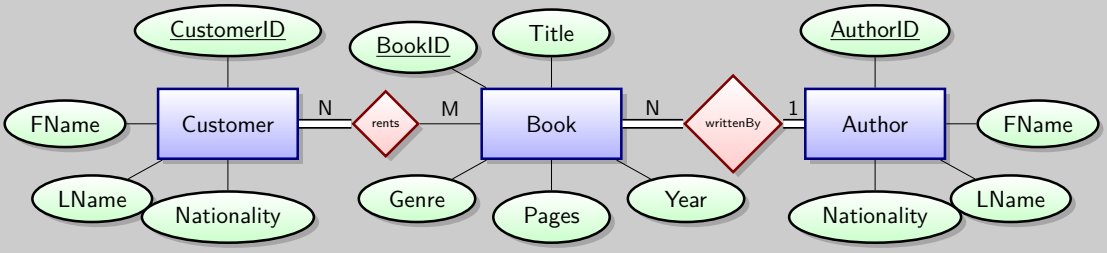
(a) Give an example for a N:M, a 1:N and a 1:1 relationship.

Solution:

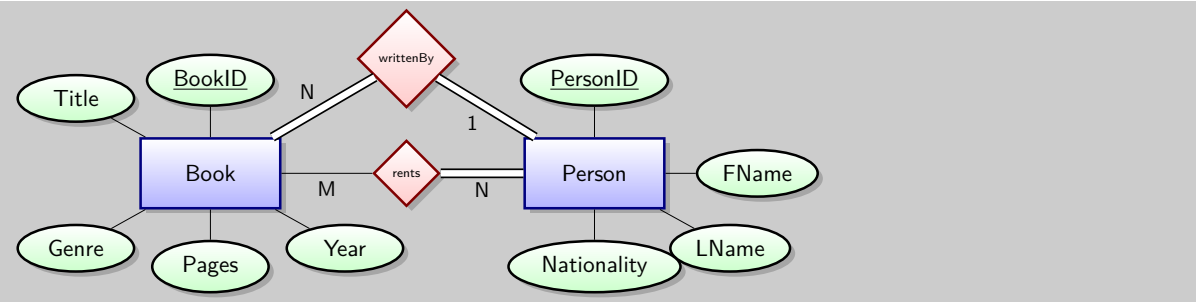


(b) Create an Entity Relationship diagram for a library where you want to store information about the customers, books and authors. Think about useful attributes and relationships.

Solution:



If we assume that authors of books are also customers in the library, the ER diagram should look like this:


writtenBy
BookID
Title
N
1
PersonID
Book
rents
Person
FName
M
N
Genre
Pages
Year
Nationality
LName

The following questions are only for the 10 ETCS students

1. **Understanding python code** For each of the following Python functions, describe what the algorithm does and how it works and provide the output for the given inputs.

   (a)
   ```
   def function1(lst):
       x = lst[0]
       for num in lst:
           if num < x:
               x = num
       return x

   # What is the result of
   function1([3, 1, 4, 1, 5, 9, 2, 6, 5])
   ```

   Solution: The function returns the smallest number in the list. The algorithm works by iterating over the list and updating the smallest number found so far.
   `function1([3, 1, 4, 1, 5, 9, 2, 6, 5]) = 1`

   (b)
   ```
   def function2(lst, e):
       c = 0
       for i in lst:
           if i == e:
               c += 1
       return c

   # What is the result of
   function2([1, 2, 2, 3, 2, 4],2)
   ```

   Solution: The function returns the number of occurrences of the element e in the list. The algorithm works by iterating over the list and counting the number of times the element e appears.
   `function2([1, 2, 2, 3, 2, 4],2) = 3`

2. **K-Means Algorithm**

   (a) Explain the basic idea of the **K-means** algorithm
   (sorry, that was a mistake, we were looking for the K-means algorithm and not K-NN, this will not happen in the exam :))

   Solution: The K-means algorithm is a clustering algorithm that is used to group similar data points into clusters. It works by iteratively updating the cluster centroids until the clusters converge. The algorithm starts by randomly selecting K cluster centroids. Then, it assigns each data point to the nearest centroid and updates the centroids to the mean of the data points assigned to them. This process is repeated until the centroids no longer change significantly.

   (b) What happens if k is small (k= 1) or large (k=n)?

Solution: If k=1, the k-means algorithm will place all data points into a single cluster, failing to capture any meaningful structure within the data. This results in a high loss function value, as the variance within the cluster will be maximized.

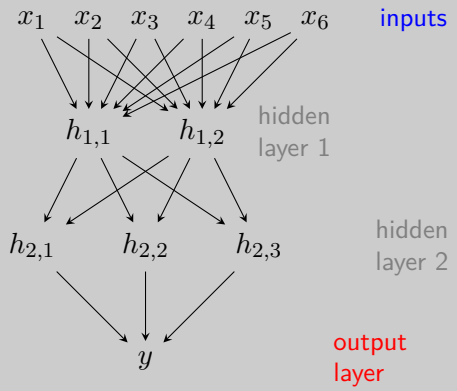
On the other hand, if k=n (where n is the number of data points), each data point will form its own cluster. In this case, the loss function will be minimized to zero, as each point is perfectly assigned to its own cluster. However, this solution lacks generalization and does not provide any useful insights into the underlying data distribution.

In both extreme cases, the k-means algorithm fails to identify meaningful patterns, emphasizing the importance of selecting an appropriate value for k, which is often done using techniques such as the elbow method.

3. **MLP**

   (a) Draw a Deep Neural Network with 6 input features, 2 hidden layers (2 and 3 neurons), and an output layer (1 neuron).

Solution:



   (b) How many parameters does the above network have?

Solution: One neuron with k inputs has $k+1$ parameters.

- hidden layer 1: $2 \times (6+1) = 14$
- hidden layer 2: $3 \times (2+1) = 9$
- output layer: $1 \times (9+1) = 10$

In total: $14 + 9 + 10 = 33$ parameters.

4. **RAID Configurations** You have 8 hard drives, each with a capacity of 10 TB, where one is used as hot spare. Your goal is to explore and compare different RAID configurations (RAID 0, RAID 1, RAID 5, and RAID 6)

   (a) For each RAID configuration (0, 1, 5, 6), calculate the **usable storage capacity**.

Solution:

- RAID 0: 7 drives, 10 TB each, 70 TB usable storage

- RAID 1: All disks are used for mirroring, therefore 10 TB usable storage
- RAID 5: From seven disks are seven used for striping data and one for storing the parity, therefore 60 TB usable storage
- RAID 6: From seven disks are 5 used for striping data and two for storing the parity, therefore 60 TB usable storage

(b) For each RAID level, describe how many disk failures the system can tolerate before data is lost.

Solution:

- RAID 0: 0 disk failure
- RAID 1: n-1 disk failures
- RAID 5: 1 disk failures
- RAID 6: 2 disk failures