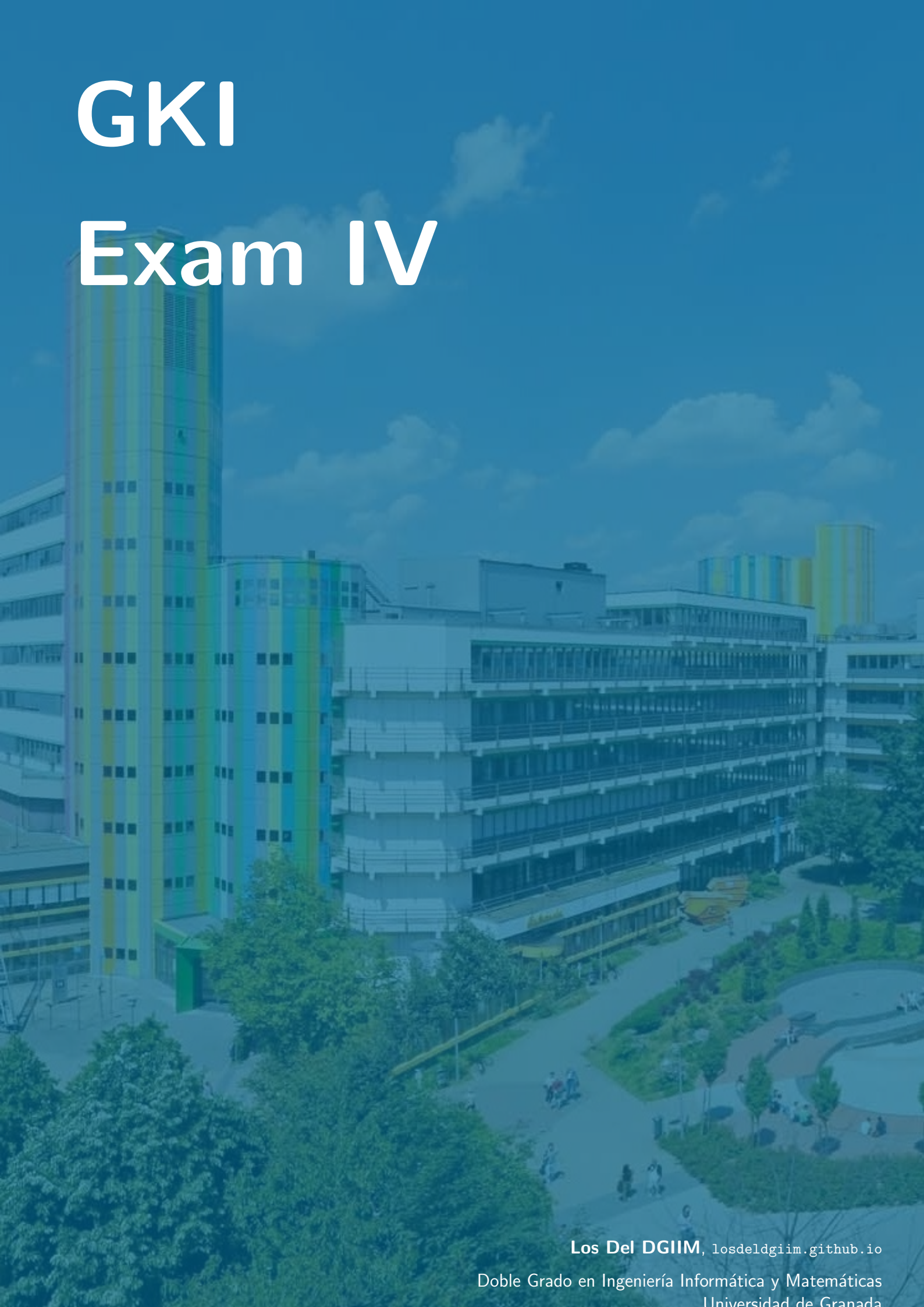
GKI

Exam IV



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas

Universidad de Granada

GKI
Exam IV

Los Del DGIIM, `losdeldgiim.github.io`

Arturo Olivares Martos

Granada, 2026

Subject Fundamentals of Artificial Intelligence.

Academic Year Summer Semester 2023.

Lecturer Gérald Kämmerer.

Description Final Exam.

Date 01.08.2023.

Duration 90 minutes.

Part 1: Short Questions (54 pts.)

Ejercicio 1 (Unsupervised Learning (3 pts.)). What is unsupervised learning? Mark whether each of the following methods belongs to unsupervised learning or not.

Method	Right	Wrong
PCA	☐	☐
Hill Climbing	☐	☐
kNN	☐	☐
kMeans	☐	☐
t-SNE	☐	☐
GANs	☐	☐

Tabla 1: Classification of Unsupervised Learning Methods

Ejercicio 2 (KDD Process (2 pts.)). What are the steps involved in the Knowledge Discovery in Databases (KDD) process?

Ejercicio 3 (Date Conversion Issues (1 pt.)). Indicate a possible problem when converting the date 08/01/23.

Ejercicio 4 (Basic Statistics (3 pts.)). Given the numbers $A = \{3, 4, 5, 10, 11, 1, 8\}$, compute the median, mean, and variance.

Ejercicio 5 (Data Aggregation Problems (2 pts.)). What problems can occur when performing data aggregation? Mark whether the statement is right or wrong.

Problem	Right	Wrong
Information gain	☐	☐
Information loss	☐	☐
Data reduction	☐	☐
Data separation	☐	☐

Tabla 2: Data Aggregation Issues

Ejercicio 6 (Curse of Dimensionality (2 pts.)). Where does the effect of the "Curse of Dimensionality" occur?

Method	Right	Wrong
High-dimensional search	☐	☐
Data aggregation	☐	☐
Optimize a neural network	☐	☐
Execute a binary classifier	☐	☐

Tabla 3: Curse of Dimensionality Contexts

Ejercicio 7 (Search Methods (1 pt.)). For which search method is Depth-Limited Search a modification?

Ejercicio 8 (Topological Search (5 pts.))**.** Give an example to explain Topological Search.

Ejercicio 9 (Genetic Algorithms - Mutation & Crossover (3 pts.))**.** Given two genes of length 10 with bases $\{0, 1, 2\}$:

- Gene 1: 0101211221
- Gene 2: 2111111001

Apply mutation and a crossover type to these genes starting with Gene 1.

Ejercicio 10 (Uniform Cost Search (4 pts.))**.** Explain Uniform Cost Search using the graph in Figure 1. Go through the different steps.

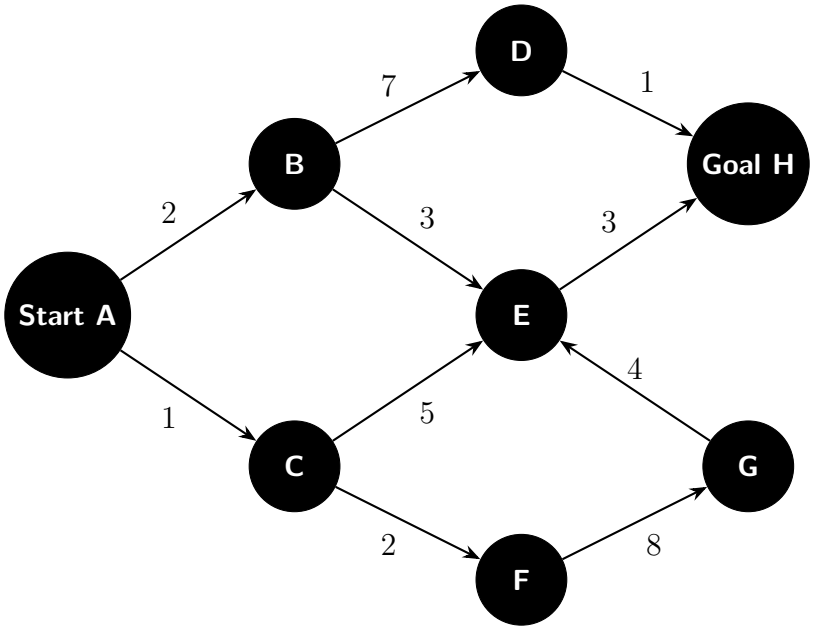


Figura 1: Uniform Cost Graph

Ejercicio 11 (Roulette Selection (2 pts.))**.** What is Roulette Selection in genetic algorithms?

Ejercicio 12 (Underfitting vs. Overfitting (2 pts.))**.** Explain what is meant by underfitting and overfitting in machine learning.

Ejercicio 13 (Confusion Matrix Elements (6 pts.))**.** Fill in the missing entries correctly in Table 4 and explain the elements of the confusion matrix.

Predicted / Actual	**Actual Yes**	**Actual No**	**Actual Total**
Predicted Yes	45 %	15 %	
Predicted No		15 %	40 %
Predicted Total	70 %		

Tabla 4: Confusion Matrix

Ejercicio 14 (F1-Score Definition (1 pt.))**.** How is the F1-Score defined?

Ejercicio 15 (Resampling Methods (2 pts.))**.** Name one resampling method and explain how it works.

Ejercicio 16 (Linear vs. Ridge vs. Lasso (3 pts.))**.** List the differences between Linear Regression, Ridge Regression, and Lasso Regression.

Ejercicio 17 (Principal Component Analysis (1 pt.))**.** What happens during Principal Component Analysis (PCA)?

Ejercicio 18 (Top-Down Approach (2 pts.))**.** What is the top-down approach in hierarchical clustering or decision trees?

Ejercicio 19 (Linkage Types in Hierarchical Clustering (4 pts.))**.** Name two types of linkage in hierarchical clustering and state the mutual advantages and disadvantages of each.

Ejercicio 20 (Activation Function Definition (1 pt.))**.** Name an activation function and give its mathematical definition.

Ejercicio 21 (Feedback Connections in Neural Networks (3 pts.))**.** Draw and label a lateral feedback connection, a direct feedback connection, and an indirect feedback connection in the network given in Figure 2.

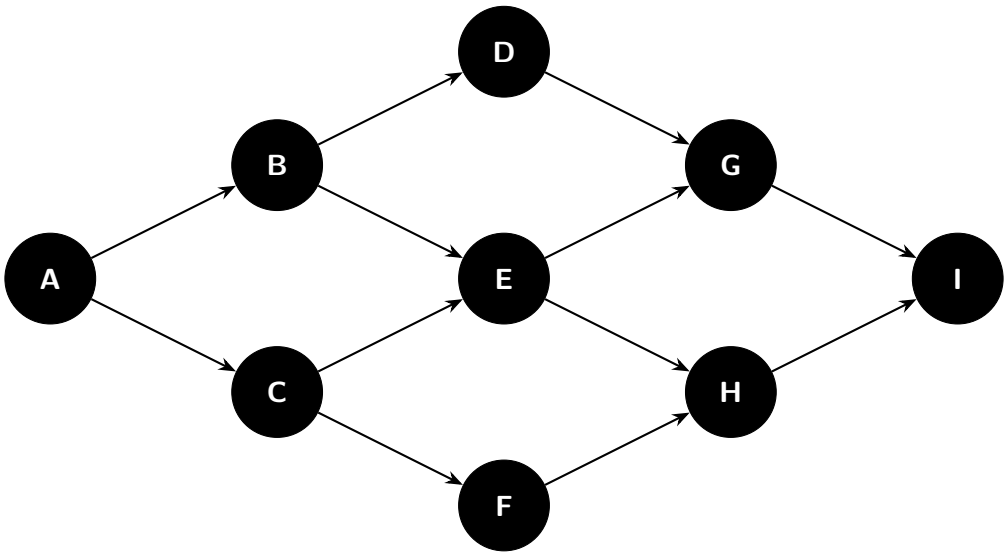


Figura 2: Forward-Feeding Network

Ejercicio 22 (Stemming (1 pt.))**.** What is stemming in Natural Language Processing (NLP)?

Part 2: kNN Classification (15 pts.)

The precipitation dataset (Table 5) is provided. The goal is to classify the two days from Table 6 as rainy or non-rainy days.

Day	Max. Temperature	Cloudiness	Rain Level	Rainy Day
1	20,1°C	30 %	0.30 mm	Yes
2	29,8°C	80 %	0.28 mm	Yes
3	21,2°C	25 %	0.00 mm	No
4	25,0°C	20 %	0.01 mm	No
5	22,5°C	85 %	0.35 mm	Yes
6	19,8°C	30 %	0.05 mm	No
7	20,2°C	80 %	0.31 mm	Yes

Tabla 5: Weather Data for kNN Classification

Ejercicio 23 (Feature Types (2 pts.))**.** Which features in Table 5 are categorical and which are nominal?

Ejercicio 24 (Distance Metrics (5 pts.))**.** Give two different distance metrics and provide their mathematical formulas. Explain the formulas and the differences between them.

Ejercicio 25 (kNN Execution (8 pts.))**.** Classify the two days given in Table 6 using the kNN algorithm with $k = 4$.

Day	Max. Temperature	Cloudiness	Rain Level
1	21,3°C	45 %	0.25 mm
2	18,3°C	25 %	0.00 mm

Tabla 6: Data to Classify

Hint: Use the following modified distance metric:

$$d(\text{Day 1}, \text{Day 2}) = |\text{Difference in Cloudiness (\%)}| + 100 \cdot |\text{Difference in Rain Level (mm)}|$$

Example:

$$d(\text{Day 1}, \text{Day 2}) = |30 - 80| + 100 \cdot |0{,}30 - 0{,}28| = 50 + 100 \cdot 0{,}02 = 50 + 2 = 52$$

Specify the distance ranking explicitly for both days and state the final classification decision.

Part 3: Decision Tree & Random Forest (20 pts.)

Ejercicio 26 (ID3 Algorithm Steps (4 pts.))**.** Name and explain the steps of the ID3 algorithm.

Ejercicio 27 (ID3 vs. C4.5 (2 pts.))**.** Name one difference between the ID3 and C4.5 algorithms.

Ejercicio 28 (Gini Impurity Calculation (3 pts.))**.** Consider the dataset with 20 points given in Table 7 and illustrated in Figure 3. Features F_1 and F_2 correspond to $Feature1$ and $Feature2$. Calculate the individual Gini impurities and the total Gini impurity for the split at $F_2 = 4$ based on the target class *Type* (Pear/Apple).

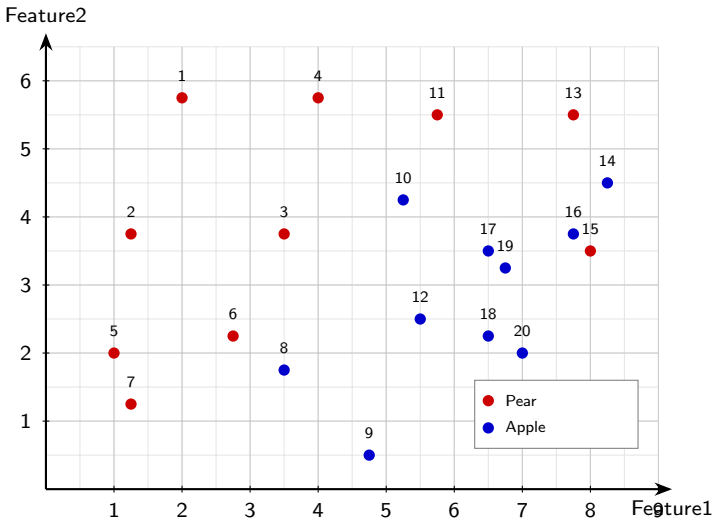


Figura 3: Scatter Plot of Feature1 vs Feature2

Ejercicio 29 (Decision Tree Construction (5 pts.))**.** Table 8 provides several candidate splits and their corresponding Gini Impurities. Create a Decision Tree that correctly classifies all points based on this table. Draw it explicitly and specify its depth.

Ejercicio 30 (Bootstrap Samples (1 pt.))**.** Why are bootstrap samples used when building a Random Forest?

Ejercicio 31 (Random Forest vs. Decision Tree Classification (5 pts.))**.** Classify the point ($F_1 = 4, F_2 = 5{,}25$) using the Random Forest in Figure 3. State the prediction of each individual tree as well as the aggregated output. Then, classify the same point using the large Decision Tree on the right (draw the decision path).

Index	F_1	F_2	**Type**
1	2.00	5.75	Pear
2	1.25	3.75	Pear
3	3.50	3.75	Pear
4	4.00	5.75	Pear
5	1.00	2.00	Pear
6	2.75	2.25	Pear
7	1.25	1.25	Pear
8	3.50	1.75	Apple
9	4.75	0.50	Apple
10	5.25	4.25	Apple
11	5.75	5.50	Pear
12	5.50	2.50	Apple
13	7.75	5.50	Pear
14	8.25	4.50	Apple
15	8.00	3.50	Pear
16	7.75	3.75	Apple
17	6.50	3.50	Apple
18	6.50	2.25	Apple
19	6.75	3.25	Apple
20	7.00	2.00	Apple

Tabla 7: Fruit Dataset

Feature	**Split**	**Gini Impurity**
F_1	< 1	0.487
F_1	≤ 2	0.469
F_1	< 3	0.313
F_1	< 4	0.444
F_1	< 5	0.374
F_1	< 6	0.487
F_1	≤ 7	0.500
F_1	< 8	0.421
F_2	< 1	0.498
F_2	< 2	0.493
F_2	< 3	0.479
F_2	< 4	0.493
F_2	< 5	0.375

Tabla 8: Feature Splits with Gini Impurity

Part 4: Encoding & Forward Propagation (5 pts.)

Ejercicio 32 (Ordinal Encoding (2 pts.))**.** Given a table column with the elements {"OH", "OR", "FI", "NY", "WY", "WI"}, perform ordinal encoding on these expressions and present the output in a table. What happens if you now want to add "CA" to the list?

Ejercicio 33 (Forward Propagation Network (2 pts.))**.** Given the small neural net-

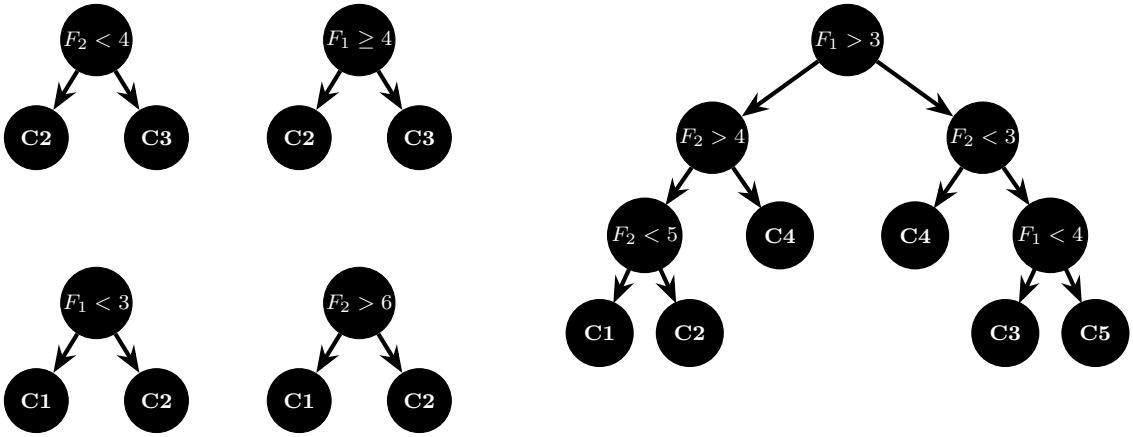


Figura 4: Random Forest & Decision Tree

work below with input $x = 12$, calculate the output values after applying the activation functions and write them into the corresponding white boxes.

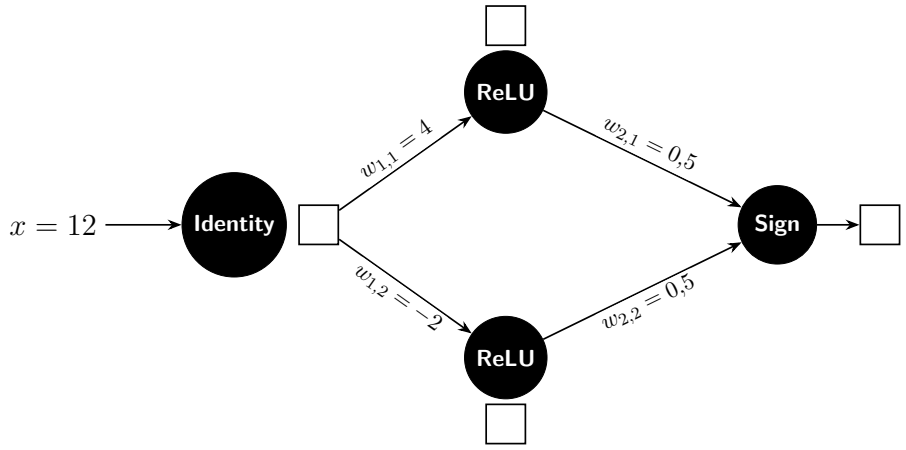


Figura 5: Small Feedforward Neural Network

Ejercicio 34 (Error Calculation - MAE & MSE (1 pt.))**.** Suppose the neural network outputs 2, and the target value should be 4. Calculate the Mean Absolute Error (MAE) and Mean Squared Error (MSE).