OSDA HW2

Tokkozhin Arsen

October 2023

Problem 1: Prove Isomorphism and Determine Topological Sorting (3 points)
Given:

- Set A = {1,2,3,4,5,6} with relation R_1 such that (a,b) ∈ R_1 if and only if a + b is even
- Set B = {'a', 'b', 'c', 'd', 'e', 'f'} with a relation $R`_2$ such that (x,y) ∈ $R`_2$ if and only if the alphabetical distance between x and y is even.

Tasks:

1. Prove that the relations R_1 and $R`_2$ are isomorphic.
2. Provide one possible topological sorting for the graph of R_1 or $R`_2$

Solution
To prove that two relations are isomorphic, we need to show that there exists a bijective function f between the sets A and B that preserves the relation properties. In other words, we need to find a bijection f: A → B such that for any (a, b) ∈ R1, (f(a), f(b)) ∈ R'2, and conversely.
We can map the elements in A to the elements in B based on their positions in the sets:

- f(1) = a
- f(2) = b
- f(3) = c
- f(4) = d
- f(5) = e
- f(6) = f

In other words we can define f as function for every element a from set A, f(a) = b, where b is letter from set B, which alphabetical number is equal to a.
Lets prove that f is bijective function:

- For any (a, b) $\in$ R1, if a + b is even, then the alphabetical distance between f(a) and f(b) will also be even because the mapping preserves the order.
- Similarly, for any (x, y) $\in$ R'2, if the alphabetical distance between x and y is even, then a + b will also be even because the mapping preserves the order.

Overall out f will be a bijective function because it is both surjective and injective. Since the function f satisfies the conditions for both R1 and R'2, we can conclude that the relations R1 and R'2 are isomorphic.
Topological sorting is typically applied to directed acyclic graphs (DAGs), and it may not always be possible to create a topological sorting for all relations. In the case of the relations R1 and R'2, they are not directed acyclic graphs (DAGs) because they contain cycles. For instance R1 has cycle:

$$(2,4) \rightarrow (4,6) \rightarrow (6,2)$$

and R'2 has cycle too:

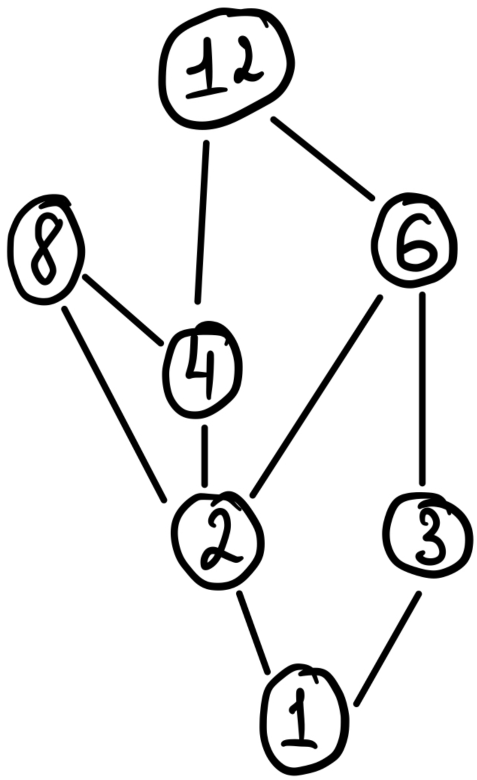
$$(b,d) \rightarrow (d,f) \rightarrow (f,b)$$

Therefore, it is not possible to create a topological sorting for these relations.

Problem 2: On Posets Properties (3 points)
Consider the set S = {1,2,3,4,6,8,12}. Define a relation $\leq$ on S such that for any two numbers a and b in S, a $\leq$ b if and only if a divides b.

1. Draw the Hasse Diagram representing the poset (S, $\leq$)
2. List three distinct antichains from the poset.
3. Define three order ideals from this poset.
4. Provide three order filters from the poset.

Solution
Lets draw the Hasse Diagram representing the poset (S, $\leq$):



Three distinct antichains from the poset:

1. (2,3)
2. (3,4)
3. (3,8)

Let $(P, \leq)$ be a poset. A subset $J \subseteq P$ is called an order ideal if for any $x \in J$ and $y \leq x$, it implies that $y \in J$.
Three order ideals from the poset:

1. (1,2,4,8)
2. (1,3,6,12)
3. (1,2,4,12)

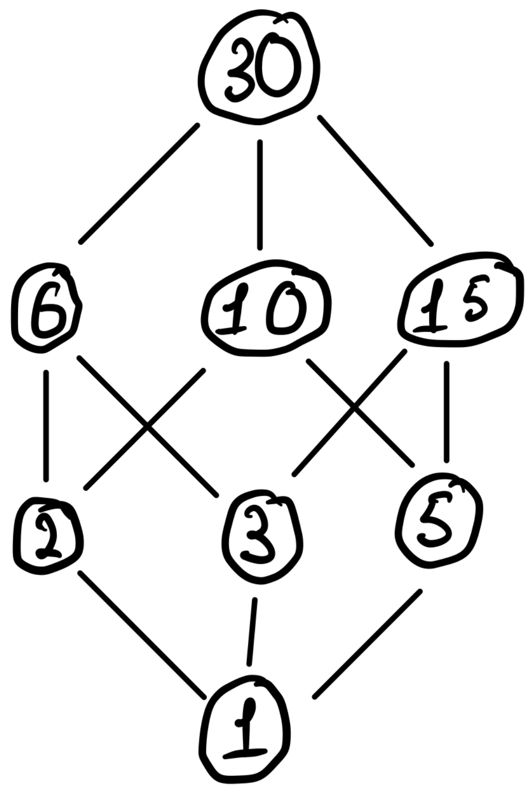
Similary a subset $F \subseteq P$ is called an order filter if for any $x \in F$ and $y \geq x$, it implies that $y \in F$.
Three order filters from the poset:

1. (8)
2. (12)
3. (2,4,6,8,12)

Problem 3: On Order Dimension (2 points)
Consider the set P = $\{1, 2, 3, 5, 6, 10, 15, 30\}$. Define a relation $\leq$ on P such that for any two numbers a and b in P, a $\leq$ b if and only if a divides b.

1. Draw the Hasse diagram representing the poset (P, $\leq$)
2. Determine the order dimension of the poset (P, $\leq$). Justify your answer.

Solution
The Hasse Diagram representing the poset (P,$\leq$):



Lets determine the order dimension of the poset (P, $\leq$).
As we remember the order dimension of a poset is the minimum count of linear orderings where their intersection gives the partial order in consideration. Consider that this order dimension of poset P, is Od(p).
It is obvious that Od(P) > 1, because (P, $\leq$) dont show complete order.
Consider Od(P) = 2 and there is a two orders: A and B, A $\cap$ B is (P, $\leq$).
A:

$$a_1 \leq a_2 \leq a_3 \leq a_4 \leq a_5 \leq a_6 \leq a_7 \leq a_8$$

B:

$$b_1 \leq b_2 \leq b_3 \leq b_4 \leq b_5 \leq b_6 \leq b_7 \leq b_8$$

It is clearly that $a_1 = b_1 = 1$ and $a_8 = b_8 = 30$
From order A we can say that for other elements:

$$2 \leq 3 \leq 5$$

$$6 \leq 10 \leq 15$$

To achieve the condition that A ∩ B is (P, ≤), for B it must be:

$$5 \leq 3 \leq 2$$

$$15 \leq 10 \leq 6$$

Which is not possible, because order of this elements in B is inversion of their order in A.
So lets check situation when Od(P) = 3.
There are 3 linear orders: A:

$$1 \leq 2 \leq 3 \leq 5 \leq 6 \leq 10 \leq 15 \leq 30$$

B:

$$1 \leq 3 \leq 5 \leq 2 \leq 10 \leq 15 \leq 6 \leq 30$$

C:

$$1 \leq 5 \leq 2 \leq 3 \leq 15 \leq 6 \leq 10 \leq 30$$

where A ∩ B ∩ C = (P, ≤) Answer: 3 linear orders.

Problem 4: Big homework (2 points)
Perform classification on the selected data using standard ML tools:

- Decision tree
- Random forest
- xGboost
- Catboost
- k-NN
- Naive Bayes
- logistic regression

Use Cross-validation (at least 5 fold) to tune the model parameters and accuracy and f1-score to assess the models.
Solution
In every pipeline for each dataset i used Optuna framework for hyperparameters tuning + 5 fold CV.
Table with results:
Water-potability dataset:

Model	**Accuracy**	**F1**
Decision Tree	0.6250	0.5079
Random Forest	0.6280	0.3858
XGBoost	0.6631	0.6105
CatBoost	0.6662	0.6120
KNN	0.5320	0.4825
Naive Bayes	0.6311	0.5268
Logistic Regression	0.6280	0.3858

Table 1: Water-potability classification

Heart attack dataset:

Model	**Accuracy**	**F1**
Decision Tree	0.8361	0.8360
Random Forest	0.8525	0.8510
XGBoost	0.8361	0.8360
CatBoost	0.8689	0.8685
KNN	0.6557	0.6497
Naive Bayes	0.8852	0.8847
Logistic Regression	0.8197	0.8195

Table 2: Heart attack classification

Healthcare diabetes dataset:

Model	**Accuracy**	**F1**
Decision Tree	0.7115	0.6901
Random Forest	0.7628	0.6963
XGBoost	0.7115	0.6800
CatBoost	0.7179	0.6796
KNN	0.7179	0.6565
Naive Bayes	0.7372	0.7029
Logistic Regression	0.7372	0.6933

Table 3: Healthcare diabetes classification