Sets and Functions Exercises (Pre-Reading for Data Science MSc)

1. Define a set and give two examples relevant to datasets in data science.

Set = a defined collection of distinct objects
Eg. set of features (income, age) or
set of classes (spam, not spam)

2. State the difference between membership ($\in$) and non-membership ($\notin$).

$x \in A \rightarrow x$ belongs to A
$x \notin A \rightarrow x \neg$ belong to A

3. Let $A = \{1, 2, 3\}, B = \{2, 3, 4\}$. Compute $A \cup B$ and $A \cap B$.

$A \cup B = 1, 2, 3, 4$
$A \cap B = 2, 3$

4. Compute $A \setminus B$ and $B \setminus A$ for the sets above.

$A \setminus B = 1$
$B \setminus A = 4$

5. Verify De Morgan's law for sets $A = \{1, 2\}, B = \{2, 3\}$ with universal set $U = \{1, 2, 3\}$.

De Morgan's Law $\rightarrow \neg(P \wedge Q) \equiv (\neg P \vee \neg Q)$
$\neg(P \vee Q) \equiv (\neg P \wedge \neg Q)$

$\neg(A \wedge B) \equiv (\neg A \vee \neg B)$
$= \neg(2) \equiv (\neg B \vee 1)$
$= 3, 1 \equiv 3, 1$

6. List all subsets of $A = \{x, y\}$ and compute $\mathcal{P}(A)$.

Subsets = $\emptyset, \{x\}, \{y\}, \{x, y\}$

$P(A) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$

7. State the cardinality of $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ (finite or infinite, and type of infinity).

Cardinality → No. of elements in a set, $|A|$

$\mathbb{N} \rightarrow \infty$ - countably
$\mathbb{Q} \rightarrow \infty$ - countably
$\mathbb{R} \rightarrow \infty$ - uncountably

8. For relation $R = \{(a, b) : a \leq b, a, b \in \{1, 2, 3\}\}$, check reflexivity.

$$\left.\begin{array}{l} 1 \leq 1 \\ 2 \leq 2 \\ 3 \leq 3 \end{array}\right\} \therefore R \text{ is reflexive}$$

9. For the same relation, check symmetry and antisymmetry.

Symmetry → if $a \leq b$, $b \geq a$
↳ false → $R \neg$ symmetric

Antisymmetric → $a \leq b$: $1 \leq 2$
$b \leq c$: $2 \leq 3$
$a \leq c$: $1 \leq 3$

$\therefore$ R is antisymmetric.

10. Is $R = \{(a, b) : a - b \text{ is even}, a, b \in \mathbb{Z}\}$ an equivalence relation?

Reflexive $\rightarrow$ for any $\mathbb{Z}$, $a - a \equiv$ even

Symmetric $\rightarrow$ $(a, b) \in \mathbb{R}$ is even
$b - a = -(a - b)$
so $b, a \in \mathbb{R} \rightarrow$ even

Transitive $\rightarrow$ $a - b$ is even, $b - c$ is even
$(a - b) + (b - c) = a - c$
As the sum of 2 evens $\equiv$ even
$a - c \Rightarrow$ even

$\therefore$ R = equivalence relation

11. Partition $\mathbb{Z}$ into equivalence classes defined by $a \sim b \iff a - b$ is divisible by 3.

$a \sim b$ means a is related to b
$\Longleftrightarrow$ means if and only if

$[0] = \{\ldots, -6, -3, 0, 3, 6, \ldots\} = 0 \pmod 3$
$[1] = \{\ldots, -5, -2, 1, 4, 7, \ldots\} = 1 \pmod 3$
$[2] = \{\ldots, -4, -1, 2, 5, 8, \ldots\} = 2 \pmod 3$

12. Define domain, codomain, and range of function $f(x) = \frac{1}{x}$.

Domain = $\mathbb{R} \setminus \{0\}$
Codomain = $\mathbb{R}$
Range = $\mathbb{R} \setminus \{0\}$

13. For $f(x) = 2x + 1$, state if $f : \mathbb{Z} \to \mathbb{Z}$ is injective, surjective, bijective.

Injective if $f(a) = f(b) \Rightarrow a = b$
(If outputs are equal, inputs must be equal)

$2a + 1 = 2b + 1 \Rightarrow a = b$

Surjective if every element of codomain $\mathbb{Z}$
has some preimage in the domain

$2x + 1$ always hits an odd number
↳ no even is ever hit

$\therefore f(x) = 2x + 1$ is surjective only.

14. For $f(x) = x^2$ with domain $\mathbb{R}$ and codomain $\mathbb{R}$, find its range.

Range $= [0, \infty]$

15. Let $f(x) = x + 1, g(x) = 2x$. Compute $(f \circ g)(x)$ and $(g \circ f)(x)$.

$(f \circ g)(x) = (f(g(x)) = 2x + 1$
$(g \circ f)(x) = (g(f(x)) = 2x + 2$

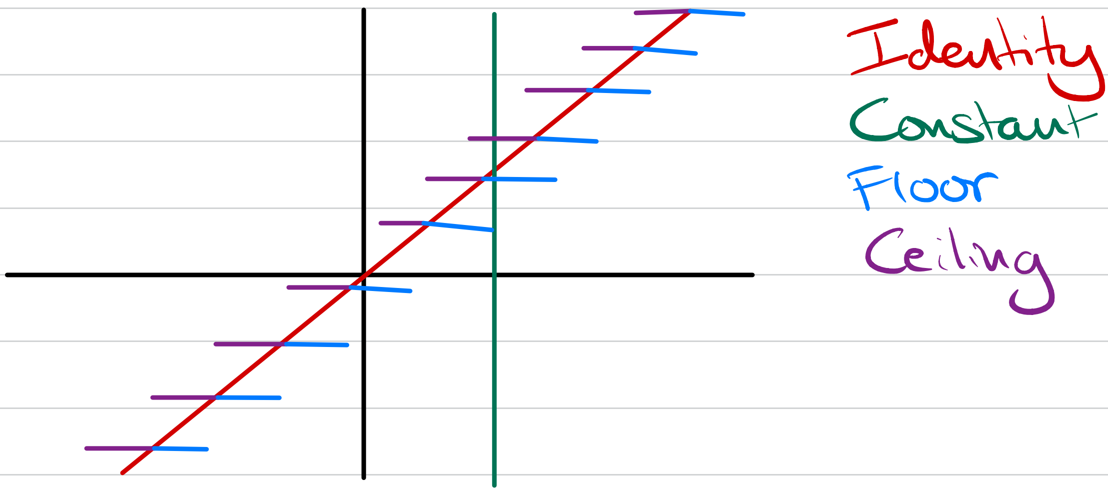
16. Find the inverse of $f(x) = 3x + 4$ if it exists.

$y = 3x + 4$
$3x = y - 4$
$x = \dfrac{y-4}{3}$

$\therefore f^{-1}(x) = \dfrac{x-4}{3}$

Made with Goodnotes

17. Sketch the graphs of identity, constant, floor, and ceiling functions.



18. Provide a data science example where injectivity matters (e.g. ID mapping).

Unique user ID $\mapsto$ one record (no collisions)

19. Provide a data science example where surjectivity matters (e.g. coverage of categories).

Category encoding – must hit all categories

20. Explain how equivalence relations link to clustering in machine learning.

Clustering → equivalence classes group data points with same label/cluster