

## Combinatorics Exercises (Pre-Reading for Data Science MSc)

1. State the rule of sum and give a dataset-related example.

Rule of sum:

- If A has  $m$  ways, task B has  $n$  ways, total  $m+n$
- E.g., choose model (3 options) OR preprocessing method (2 options)  $\rightarrow 5$  total

2. State the rule of product and give an example of feature combinations.

Rule of product:

- If A has  $m$  ways, B has  $n$  ways, multiply  $m, n$
- E.g., choose model (3) AND preprocessing (2)  $\rightarrow 6$  total

3. Compute number of possible 4-digit PINs using digits 0-9 (repetition allowed).

$$10^4 = 10,000$$

4. Compute number of ways to order 6 books on a shelf.

$$6! = 720$$

5. Compute number of unique permutations of word "LEVEL".

$$\frac{n!}{n_1 n_2 \dots} = \frac{5!}{2! 2!} = 30$$

6. Compute number of unique permutations of word "BALLOON".

$$\frac{7!}{2!2!} = 1260$$

7. From 12 students, choose 4 for a group project. How many ways?

$$\frac{n!}{r!(n-r)!} = \frac{12!}{4!(8)!} = 495$$

8. From 8 features, choose 3 to include in a model. How many subsets?

$$\frac{8!}{3!5!} = 56$$

9. A dataset has 5 categories. Choose 3 categories (allowing repeats). How many ways?

$$\binom{5+3-1}{2} = \frac{7!}{3!4!} = 35$$

10. How many binary strings of length 5 exist?

$$2^5 = 32$$

11. Expand  $(x+y)^4$  using binomial theorem.

$$1x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + 1y^4$$

12. Expand  $(0.3x + 0.7y)^3$  (probabilistic weights).

$$0.027x^3 + 0.189x^2y + 0.441xy^2 + 0.343y^3$$

13. Use Pascal's triangle to find coefficients of  $(x + y)^5$ .

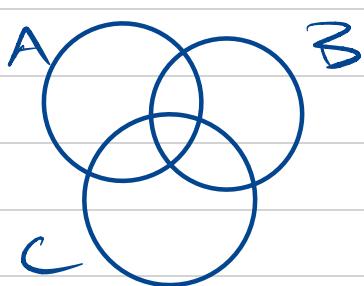
$$\begin{array}{ccccccc}
 & & 1 & & & & \\
 & & | & | & | & & \\
 & 1 & & 2 & & 1 & \\
 & | & & 3 & & 3 & | \\
 & 1 & & 4 & & 6 & 4 & 1 \\
 & | & & 5 & & 10 & 10 & 5 & 1
 \end{array}$$

$\therefore \text{ans} = 1, 5, 10, 10, 5, 1$

14. In a class, 20 study maths, 15 study CS, 10 study both. How many study at least one?

$$20 + (15 - 10) = 25$$

15. Generalise inclusion-exclusion for 3 sets.



$$\begin{aligned}
 |A \cup B \cup C| = & \\
 |A| + |B| + |C| - & |A \cap C| - |A \cap B| - |B \cap C| \\
 - & |A \cap B \cap C|
 \end{aligned}$$

16. In hashing, explain how pigeonhole principle implies collisions.

If you put  $> n$  objects in  $n$  boxes, at least one box contains  $> 2$  objects

↳ These cases are collisions

↳ Collisions guaranteed if objects  $>$  boxes

Means: minimum maximum



17. If 25 pigeons in 10 holes, minimum in one hole?

$$\lceil 25 / 10 \rceil = 3$$

18. If 1000 users map into 365 days, explain collision likelihood (birthday paradox hint).

$$CL = 1 \quad \text{as} \quad 1000 > 365$$

19. Provide a data science example of using combinations (feature selection).

Feature selection  $\rightarrow$  choose  $k$  features from  $n$  total - uses combinations

20. Provide a data science example of using permutations (sequence modelling).

Sequence modelling  $\rightarrow$  different orderings of the same tokens gives you different outcomes - uses permutations