Important Mathematical Formulas with Examples

1. Mean (Average)

Formula: Mean = (Sum of all values) / (Number of values)

Examples:

- [10, 20, 30] → Mean = 60 / 3 = 20
- [4, 6, 8, 10] → Mean = 28 / 4 = 7
- [100, 200, 300, 400, 500] → Mean = 1500 / 5 = 300

2. Median

If odd count: Middle value. If even: Average of two middle values.

Examples:

- [3, 5, 7] → Median = 5
- [4, 6, 8, 10] → Median = (6+8)/2 = 7
- [10, 20, 15] → Sorted: [10, 15, 20] → Median = 15

3. Mode

Most frequent number.

Examples:

- [2, 2, 3, 4, 5] → Mode = 2
- [1, 1, 2, 3, 3, 3] → Mode = 3
- [4, 5, 6, 7] → No mode

4. Standard Deviation

Formula: $\sigma = \text{sqrt}[\ \Sigma(xi - \text{mean})^2 / N\ ]$

Examples:

- [2, 4, 4, 4, 5, 5, 7, 9] → Mean = 5, Variance = 4 → SD = 2
- [10, 10, 10, 10] → SD = 0
- [5, 10, 15] → Mean = 10, SD ≈ 4.08

5. Z-Score

Formula: Z = (x - mean) / standard deviation

Examples:

- x = 20, mean = 10, std = 5 → Z = 2
- x = 5, mean = 10, std = 5 → Z = -1
- x = 10, mean = 10, std = 5 → Z = 0

6. Term Frequency (TF)

Formula: TF = (word count in doc) / (total words)

Examples:

- 'apple apple orange' → TF(apple) = 2/3 ≈ 0.66
- TF(orange) = 1/3 ≈ 0.33
- TF(banana) = 0

7. Inverse Document Frequency (IDF)

Formula: IDF = log(Total docs / Docs with word)

Examples:

- Word in 1 of 5 → IDF = log(5/1) ≈ 0.699
- Word in all 5 → IDF = log(5/5) = 0
- Word in 2 of 5 → IDF ≈ 0.397

8. TF-IDF

Formula: TF-IDF = TF × IDF

Examples:

- TF = 0.5, IDF = 0.69 → TF-IDF = 0.345
- TF = 0.2, IDF = 0.4 → TF-IDF = 0.08
- TF = 0.1, IDF = 0.0 → TF-IDF = 0

9. Accuracy

Formula: Accuracy = (TP + TN) / (TP + TN + FP + FN)

Examples:

- TP=5, TN=3, FP=1, FN=1 → Accuracy = 8/10 = 80%
- 45 correct out of 50 → Accuracy = 90%
- 10 correct out of 20 → Accuracy = 50%

10. Precision

Formula: Precision = TP / (TP + FP)

Examples:

- TP=5, FP=1 → Precision = 5/6 ≈ 83.3%
- TP=9, FP=1 → Precision = 9/10 = 90%
- TP=2, FP=3 → Precision = 2/5 = 40%

11. Recall

Formula: Recall = TP / (TP + FN)

Examples:

- TP=5, FN=1 → Recall = 5/6 ≈ 83.3%
- TP=8, FN=2 → Recall = 8/10 = 80%
- TP=3, FN=1 → Recall = 3/4 = 75%

12. F1 Score

Formula: F1 = 2 × (Precision × Recall) / (Precision + Recall)

Examples:

- P=0.833, R=0.833 → F1 ≈ 83.3%
- P=0.9, R=0.8 → F1 = 2×0.72/1.7 ≈ 84.7%
- P=0.6, R=0.75 → F1 ≈ 66.7%

13. Confusion Matrix Terms

TP: predicted YES, actually YES

TN: predicted NO, actually NO

FP: predicted YES, actually NO

FN: predicted NO, actually YES

Used to calculate Accuracy, Precision, Recall, F1.

14. Summary

- Use Mean, Median, Mode to describe data.
- Use SD, Z-Score to understand spread.
- Use TF, IDF, TF-IDF for text analysis.
- Use Accuracy, Precision, Recall, F1 to evaluate models.