**Fundamental Concepts:**

A black and white text on a black background

Description automatically generatedA screenshot of a black background with white text

Description automatically generated

A black background with white text

Description automatically generated

**Similarity Calculations:**A screenshot of a black and white math

Description automatically generatedA screenshot of a black and white page

Description automatically generatedA black background with white text

Description automatically generated

**Normalization:**

**Min-Max: Scaling data between a specified range (typically 0 to 1).**

A white paper with black text and numbers

Description automatically generated

**Z-Score: Scaling data based on mean and standard deviation.**

A screenshot of a math test

Description automatically generated

A diagram of a mathematical equation

Description automatically generated

**Decimal: Scaling data by shifting the decimal point.**

A screenshot of a computer

Description automatically generated

**Distance Calculation:**

Manhattan: Sum of the absolute differences between corresponding attributes' values.

manhatten = sum(abs(a1 - b1) for a1, b1 in zip(v1, v2))

**Chi-Square Test:**

Statistical test to determine if there is a significant association between categorical variables in a contingency table.

A black and white screen with white text

Description automatically generated

A screenshot of a black and white math

Description automatically generated

A screenshot of a black and white page

Description automatically generated

**A black screen with white text

Description automatically generated**

**Information Gain:**

Measure of the effectiveness of a feature in classifying a dataset.

A math equations and formulas

Description automatically generated with medium confidence

**Gini Index:**

Measure of impurity or disorder in a dataset's target variable.

A screenshot of a black and white text

Description automatically generated

**Converting Ordinal Values to Numeric:**

Assigning numerical values to ordinal categories while preserving their order.

Ordinal Encoding Techniques:

Label Encoding: Assigning each unique category a unique integer value. The assignment is typically done based on the order of appearance or alphabetical order. For instance, in the "low," "medium," "high" example, "low" might be encoded as 0, "medium" as 1, and "high" as 2.

Custom Mapping: Manually mapping ordinal categories to specific numeric values based on domain knowledge. This approach allows for more control over the mapping process, enabling the assignment of values that might reflect the inherent order or significance of the categories more accurately.

Ordinal Encoding Libraries: Some libraries in programming languages like Python (e.g., scikit-learn) offer built-in functionality for ordinal encoding. These libraries often provide methods that automatically handle the conversion of ordinal categories to numeric values while preserving the order.

Smoothing:

.A screenshot of a computer

Description automatically generatedA black and white text on a black background

Description automatically generated

Correlation from Scatter Plot:

A black and white text on a black background

Description automatically generatedA black and white text

Description automatically generated

A black and white text on a black background

Description automatically generatedA black and white text on a black background

Description automatically generated

Define TP, TN, FP, FN:

TP (True Positive): Correctly predicted positive instances.

TN (True Negative): Correctly predicted negative instances.

FP (False Positive): Incorrectly predicted as positive.

FN (False Negative): Incorrectly predicted as negative.