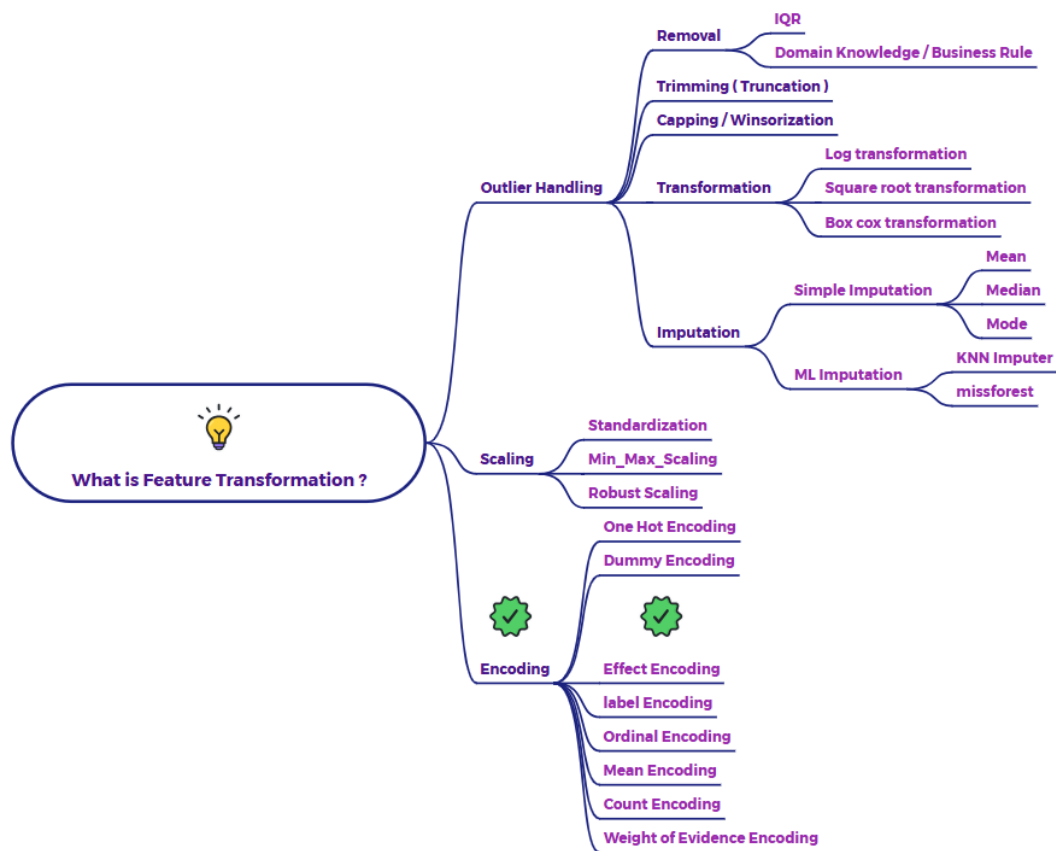


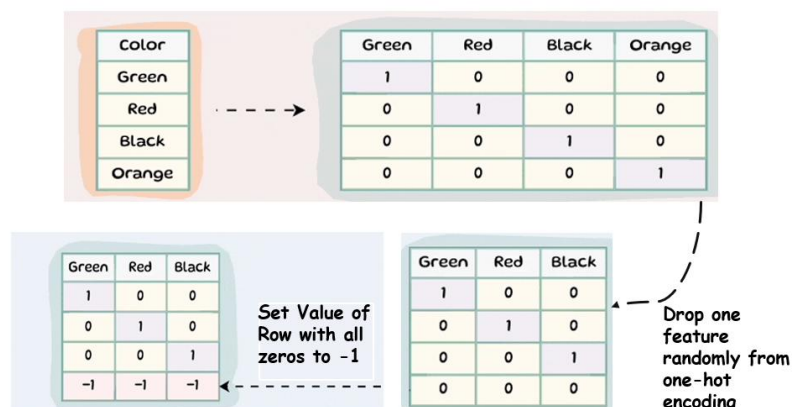
Explain Effect Encoding with an example



1. Explanation of Effect Encoding

Effect encoding is a technique used to encode categorical variables. It's similar to dummy encoding, but with a crucial difference: instead of using 0s and 1s, it uses -1, 0 and 1s. Like dummy coding, effect coding uses (n-1) columns to represent a categorical feature with n categories. The primary difference is in how the baseline category is represented. In dummy coding, the baseline category is represented by a row of zeros, whereas in effect coding, the baseline category is represented by a row of "-1"s.

Effect Encoding



2. How to Calculate Effect Encoding

Here's a step-by-step explanation with an example:

Example:

Suppose we have a dataset with a "Color" column:

Color
Green
Red
Black
Orange

1. Identify Unique Categories: The unique categories in the "Color" column are "Green," "Red," "Black," and "Orange."

2. Create Columns: Create columns for $n-1$ of the categories. For instance, we might create columns for "Color_Green", "Color_Red", and "Color_Black".

3. Encode the data:

* For the categories included in the encoding (Green, Red, Black), represent their presence with a '1' in their respective column and '0' in the other columns.

* For the baseline category ("Orange" in this case), represent it with '-1' in all of the encoded columns.

The resulting effect-encoded data looks like this:

Color	Color_Green	Color_Red	Color_Black
Green	1	0	0
Red	0	1	0
Black	0	0	1
Orange	-1	-1	-1

3. When to Use Effect Encoding

Effect encoding is particularly useful in statistical models like ANOVA (Analysis of Variance) or when you want to examine the effect of a categorical variable relative to the overall mean.

4. Strengths and Weaknesses of Effect Encoding

Strengths:

- Useful for models where you want to analyze the effects of each category relative to the overall mean of the dependent variable.
- Can be more convenient for interpreting the coefficients in some statistical models.

- **Weaknesses:**

- Slightly more complex to understand than one-hot or dummy encoding for someone new to encoding schemes.
- The -1 representation might be less intuitive for some machine learning algorithms.