assignment=40

August 1, 2023

1 Q1. What is data encoding? How is it useful in data science?

```
Ans: Data encoding is the process of converting data from one format to another, usually for th
    It is useful in data science because it enables us to work with data in a uniform and consi
```

2 Q2. What is nominal encoding? Provide an example of how you would use it in a real-world scenario.

```
Ans: Nominal encoding, also known as one-hot encoding, is a technique used in data preprocessin
```

For example, in a dataset of customer reviews, a categorical variable like "product category" could be nominal encoded into binary variables such as "is_electronics", "is_fashion", "is_home_goods", etc. This can enable us to perform analyses on the categories while preserving the non-numeric nature of the data.

3 Q3. In what situations is nominal encoding preferred over one-hot encoding? Provide a practical example.

```
Ans: There is no difference between nominal encoding and one-hot encoding. Both techniques conv
     For example, in a dataset of different car makes and models, we might use nominal encoding
```

4 Q4. Suppose you have a dataset containing categorical data with 5 unique values. Which encoding

technique would you use to transform this data into a format suitable for machine learning algorithms? Explain why you made this choice.

```
Ans: For categorical data with 5 unique values, I would use nominal encoding, also known as one
     for machine learning algorithms. This is because nominal encoding assigns a unique binary
     machine learning algorithms.
     Ordinal encoding, which assigns integer values based on the rank or order of the categorie
     order or ranking to the categories. By using nominal encoding, we can preserve the non-num
     learning algorithm to process the data effectively.
```

5 Q5. In a machine learning project, you have a dataset with 1000 rows and 5 columns. Two of the columns

are categorical, and the remaining three columns are numerical. If you were to use nominal encoding to transform the categorical data, how many new columns would be created? Show your calculations.

```
Ans: Assuming the two categorical columns have a total of k unique categories, nominal encoding
        Therefore, the number of new columns created would be:
        k * 2 (since there are two categorical columns)
    Without knowing the number of unique categories in the two categorical columns, it is not p
```

6 Q6. You are working with a dataset containing information about different types of animals, including their

species, habitat, and diet. Which encoding technique would you use to transform the categorical data into a format suitable for machine learning algorithms? Justify your answer.

```
Ans: For the given dataset containing information about different types of animals, I would use
     to transform the categorical data into a format suitable for machine learning algorithms.
     The reason for this is that nominal encoding is particularly useful when dealing with cate
     In this case, both the "species" and "habitat" variables likely do not have a natural orde
     way to convert them into binary variables that can be processed by machine learning algori
```

7 Q7.You are working on a project that involves predicting customer churn for a telecommunications

company. You have a dataset with 5 features, including the customer's gender, age, contract type, monthly charges, and tenure. Which encoding technique(s) would you use to transform the categorical data into numerical data? Provide a step-by-step explanation of how you would implement the encoding.

```
Ans: For the given dataset, the "gender" variable is nominal and the "contract type" variable i
     To transform the nominal categorical data into numerical data, I would use nominal encodin
     Here's a step-by-step explanation of how I would implement the encoding:
        1. Identify the nominal categorical variables: "gender" and "contract type"
        2. Create binary variables for each unique category in the nominal categorical variable
            we would create two binary variables, one for "Male" and one for "Female". For the
            binary variables for each type of contract (e.g., "Month-to-month", "One year", "Tw
        3. Replace the original nominal categorical variables with the newly created binary var
        4. Normalize the numerical variables (e.g., "age", "monthly charges", and "tenure") to
            processed by machine learning algorithms.
        5. Use the transformed dataset to train and evaluate machine learning models to predict
```

[ ]:

[ ]:

[ ]:

[ ]: